Studies towards the prebiotic synthesis and phosphorylation of ribonucleotides

Natalie M. Grefenstette

2017

Thesis submitted to University College London
for the Degree of Doctor of Philosophy

UCL Department of Chemistry
20 Gordon Street
London, WC1H 0AJ

Declaration of ownership

I, Natalie Marie Grefenstette, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The discovery of the universality of life's biochemistry across all domains of life suggests that life has a unified chemical origin. RNA is a highly versatile pillar of extant biochemistry, with essential and universal roles as a genetic messenger between DNA and proteins, a universally essential catalyst in ribosomes, and an essential element of a variety of co-factors. The deep-seated role of RNA in modern biochemistry and RNA's unique dual biological functionality, acting both genotypically and phenotypically, makes RNA a strong contender for the first biopolymer of life: the 'RNA world' hypothesis. However, one of the most significant problems in bridging the gap between a purely abiological and a biologically active world is a prebiotically plausible synthesis of ribonucleotides. An account of our results investigating novel prebiotically plausible syntheses of ribonucleotides is presented. We propose a new oxidative approach to the synthesis of nucleotide precursors exploiting the oxidation of acrolein, the simplest unsaturated aldehyde with many prebiotically plausible syntheses, to glycidaldehyde by aqueous hydrogen peroxide, a simple and prebiotically robust oxidant. Glycidaldehyde provides the activation required to realise regioselective incorporation of phosphate during nucleotide assembly, leading to the first prebiotically plausible aqueous synthesis of pyrimidine nucleotides 5′-phosphate. These reactions are effected under simple, mild, aqueous conditions, compatible with the rest of our prebiotic studies on the synthesis of amino acids and purine ribonucleotides. We also demonstrate that acrolein can be used in the prebiotic synthesis of C_3 amino acids, establishing acrolein as a central synthon in prebiotic chemistry. Our discovery of a generational node in the network of prebiotic chemistry that links the syntheses of amino acids with nucleotides 5'-phosphates suggests that these different groups of metabolites need not have arisen from separate chemistries, helping us to get a better understanding of the processes governing the chemical evolution of life.

Table of contents

Acknowledgements

First of all, I would like to thank my supervisor, Matt Powner, for giving me the opportunity to carry out research in a fascinating field. I would also like to thank the whole Powner Lab and previously the Motherwell group as well, for the countless hours of discussion, both chemistry-related and otherwise. Particularly, thank you to Adam, Shaun and Christian for their invaluable help in and out of the lab, as well as for their corrections and proof-reading. I would also like to thank Kathryn for the detailed and honest corrections she has given me, they have been invaluable. To all the above, and later on Becca, thank you for the support and making the last three years easier for me and more enjoyable.

Finally, a special thanks to my parents and brothers for always pushing me to do and discover more, and that from a very young age. I would not have been able to do this without the support from my friends and family. Thank you so much for your kind words, your interest and motivational speeches. To Max, I can't thank you enough for being by my side throughout everything, and for reminding me that I can do this, every single day.

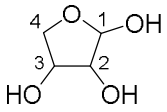
Abbreviations

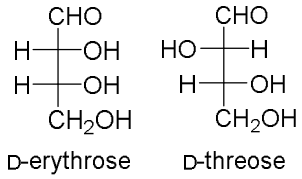
Å	angstrom
A	adenine
Ac	acetyl
Acetyl-CoA	acetyl coenzyme A
AICA	4-amino-imidazole-5-carboxamide
AICN	4-amino-imidazole-5-carbonitrile
ANA	alanyl nucleic acid
aq.	aqueous
AmTP	amidotriphosphate
ATP	adenosine triphosphate
Bz	benzoyl
C	cytosine
c	concentration
°C	degrees Celsius
calcd	calculated
cat.	catalytic
Celite®	high grade diatomaceous earth filtration agent
conc.	concentrated
COSY	correlated spectroscopy (NMR)
Δ	difference
DAMN	diaminomaleonitrile
DAP	diamidophosphate
DMAP	4-dimethylaminopyridine
DMF	*N*,*N*-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
Dowex®	ion exchange resin
eq	equivalent(s)
Et	ethyl
et al.	*et alia* (Latin: and others)
FAD	flavin adenine dinucleotide
FCC	flash column chromatography
G	guanine
GC	gas chromatography
GNA	glycol nucleic acid
h	hour(s)
HMBC	heteronuclear multiple-bond correlation
HSQC	heteronuclear single quantum correlation
Hz	Hertz
hν	electromagnetic irradiation (UV)
IR	infrared
IMP	inosine mono-phosphate
J	NMR coupling constant (measured in Hz)
KHP	potassium hydrogen phthalate
LHF	liquid hydrogen fluoride
LUCA	last universal common ancestor

M	molar
m.p.	melting point
m/z	mass/charge ratio
MCl	metal chloride
Me	methyl
MHz	megaHertz
min	minute(s)
MOH	metal hydroxide
MS	mass spectroscopy
NA	non applicable
NADH	nicotinamide adenine dinucleotide
NASA	National Aeronautics and Space Administration
NB	*Nota bene* (Latin: note well)
NMR	nuclear magnetic resonance
Ph	phenyl
P_i	inorganic phosphate, sodium salt (unless otherwise stated)
PNA	peptide nucleic acid
PP_i	pyrophosphate
ppm	parts per million
PPP_i	triphosphate
p-RNA	pyranosyl ribonucleic acid
Py	pyridine
quant.	quantititave
rac-	racemic
RNA	ribonucleic acid
RT	room temperature
sat.	saturated
soln.	solution
T	thymine
$t_{1/2}$	half life
TBAF	tetra-n-butylammonium fluoride
TBDPS	tert-butyldiphenylsilyl ether
TCC	trichloroisocyanuric acid
TEMPO	2,2,6,6-tetramethylpiperidin-1-oxyl
THF	tetrahydrofuran
TLC	thin layer chromatography
TNA	threofuranosyl nucleic acid
tr.	trace
U	uracil
UV	ultraviolet
XNA	xeno nucleic acid

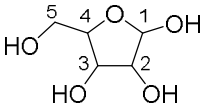
Numbering and nomenclature

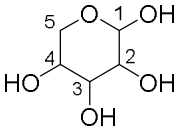
Tetrose sugars



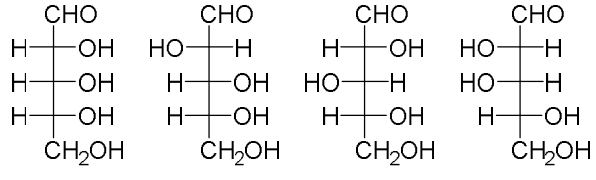


Pentose sugars

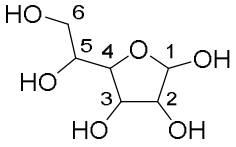




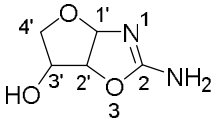
furanose pyranose



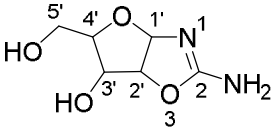
Hexose sugars

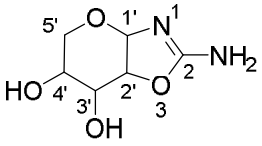


Tetrose aminooxazolines

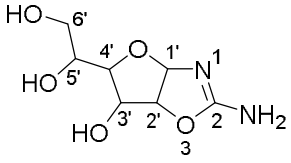


Pentose aminooxazolines

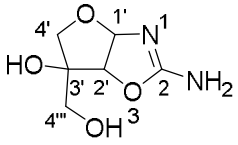




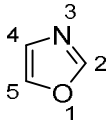
Hexose aminooxazolines



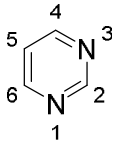
Branched aminooxazolines

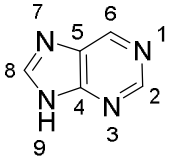


Oxazoles

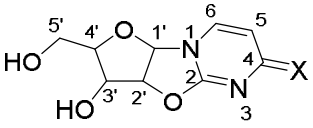


Pyrimidines (left) and purines (right)



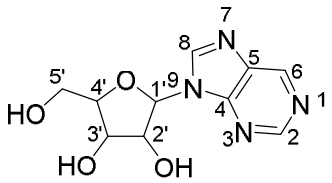


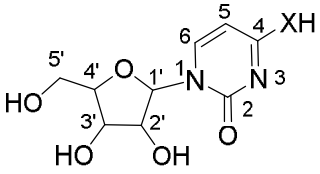
Anhydronucleosides



X= O/NH_2

Nucleosides





X= O/NH

Science is asymptotic: it never arrives at but only approaches the tantalizing goal of final knowledge. Astrology gives way to astronomy; alchemy evolves into chemistry. The science of one age becomes the mythology of the next. How will future thinkers assess our own ideas? This movement of thought – of living beings questioning themselves and their surroundings – is at the heart of the ancient question of what it means to be alive – *Lynn Margulis and Dorion Sagan*

We are coming back to spontaneous generation, albeit in a more refined and scientific sense, namely to chemical evolution – *Cyril Ponnamperuma*

If life once made itself, it must not be too difficult to make it again – *John Desmond Bernal*

Each of us carries the map of our lives on our skin, in the way we walk, even in the way we grow – *Kiran Millwood Hargrave*

1. Introduction

Understanding the origin of life, the transition from inanimate matter to a living cell, is one of the biggest unanswered questions in science. Though we may never know the precise details of the continuum of events that led to life on Earth, scientists and philosophers have, for millennia, strived to shed light on the question: "How did life start on Earth?". There are many facets to this question, and answers have been sought at all levels, from chemistry to astrophysics. Advances in biological sciences are providing an ever-increasing understanding of life, both as a whole and in molecular detail, but these advances only give vague hints to the biochemistry of the first living species. General characteristics of life can be inferred, from current biology (top down) and from chemical principles (bottom up). These suggest that a chemical system capable of replicating information has always been required in biological systems, and that metabolism and compartmentalisation from the exterior environment are essential to the function (and perhaps even the evolution) of such a biological information carrier. The fundamental characteristics of life (inheritance, growth and compartmentalisation) are not separate, but interdependent, and thus research carried out to elucidate the origins of terran life must (at least) encompass–and ultimately seek to unify–each of these fields. Understanding how these subsystems can assemble will be the cornerstone of discovering the origins of life.

1.1. Approaches to studying the origin of life

1.1.1. What is life?

> *"Life is the transmutation of energy and matter [...] Life is incessant heat-dissipating chemistry and life is memory – memory in action, as the chemical repetition of the past"*- Margulis and Sagan[1]

A fascination for understanding the rules governing our function and origin has shaped the landscape of scientific (and philosophical, historical, etc.) research. The field dedicated to this subject in its conception and name is that of biology. Tremendous amounts of knowledge have been gathered in biology in the last centuries, with the discovery of evolution, genetics and the materials behind the storage and transfer of genetic information: DNA and RNA.

Because our examples of life are limited to the one biosphere present on Earth, it is hard to make universal generalisations concerning the properties of life and how it might have originated. The theory of evolution, postulated by Charles Darwin, Alfred Russel Wallace and the physician Charles Wells, helped the scientific community realise that biology was not 'frozen' in time, but rather adapted to the environment and possibly 'emerged' from its environment[2–4]. In 1871, Darwin contemplated the possibility of life evolving out of a chemical system:

> *"But if (and oh what a big if) we could conceive in some warm little pond with all sort of ammonia and phosphoric salts, light, heat, electricity present, that a protein compound was chemically formed, ready to undergo still more complex changes"*- Darwin[5]

However, there is as yet no clear answer to the question: "What *is* life?", famously posed by Erwin Schrödinger in 1944[6]. In his titular book, he theoretically described that an "aperiodic crystal" was at the source of genetic information, a discovery that was to be made in the following decade by James Watson, Francis Crick, Maurice Wilks and Rosalind Franklin in their double-helical model of DNA[7–10] (*vide infra*). Even though these discoveries have propelled biological research to where it stands today, it in no way solves the problem of what life *is*. A panel of international experts was set up by NASA in order to devise a definition of life that could be used in the context of research into the origins of life and search for other forms of life in the universe. The definition they conceived, and that we will be basing our work on is that life is "a self-sustaining chemical system capable of Darwinian evolution"[11].

Before one delves into how life could have evolved from a chemical system, it is important to understand how the "self-sustaining chemical system" that is extant biology works. Once the structure of DNA was discovered, efforts were channelled into understanding how DNA coded for proteins. In 1954, a year after the discovery of the DNA structure was published, Gamow postulated that there was a direct link between the base sequence in DNA and the amino acid sequence in proteins[12]. He proposed that only a ≥3 letter code would generate more combinations than there are natural proteinogenic amino acids (>20 combinations to the 20 amino acids used in biology). Crick *et al.* demonstrated experimentally that amino acids in a protein sequence were most likely coded for by a group of three bases in the DNA sequence (known as a codon), that are read from a fixed starting point, do not overlap, and that the relationship between the codon and amino acids is most likely degenerate (one amino acid can be coded for by more than one codon)[13]. Once the first codon was determined by Nirenberg and Matthaei (polyuridine codes for polyphenylalanine)[14], it was followed shortly by the discovery of 54 of the 64 triplet codons[15–19], finally completed in 1968 by Khorana[20] (*Figure 1.1*).

FIRST BASE	SECOND BASE U	SECOND BASE C	SECOND BASE A	SECOND BASE G	THIRD BASE
U	Phe	**Ser**	Tyr	Cys	U
	Phe	**Ser**	Tyr	Cys	C
	Leu	**Ser**	*stop*	*stop*	A
	Leu	**Ser**	*stop*	Trp	G
C	**Leu**	**Pro**	His	**Arg**	U
	Leu	**Pro**	His	**Arg**	C
	Leu	**Pro**	Gln	**Arg**	A
	Leu	**Pro**	Gln	**Arg**	G
A	Ile	**Thr**	Asn	Ser	U
	Ile	**Thr**	Asn	Ser	C
	Ile	**Thr**	Lys	Arg	A
	Met	**Thr**	Lys	Arg	G
G	**Val**	**Ala**	Asp	**Gly**	U
	Val	**Ala**	Asp	**Gly**	C
	Val	**Ala**	Glu	**Gly**	A
	Val	**Ala**	Glu	**Gly**	G

Figure 1.1: The standard genetic code. Amino acids are represented by their standard three-letter codes. The degeneracy of the code is highlighted by the "family boxes" shown in bold.

Furthermore, it was established that the relationship between the DNA sequence and the protein sequence was indirect, and that RNA played the role of "messenger" between the two[21]. These initial ground-breaking advances in biochemistry paved the way to the central dogma of biology (*Figure 1.2*): genetic information is stored in the double helix of DNA, based on four nucleobases adenine (A), guanine (G), cytosine (C) and thymine (T), that through simple hydrogen bonding interactions (known as Watson-Crick base pairing, A=T and C≡G, *Figure 1.3*), can transfer information from one strand to the complementary strand, thereby allowing transfer of information from one generation to the next through replication[22].

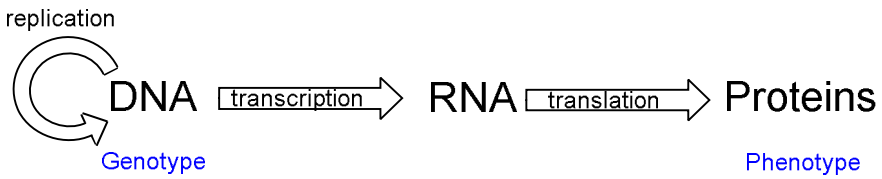


Figure 1.2: The central dogma of molecular biology.

Information in DNA is transcribed into RNA through similar hydrogen-bonding interactions (T is replaced with uracil (U) in RNA), forming messenger RNA (mRNA). This mRNA sequence is then translated into proteins thanks to the ribozyme and transfer RNA (tRNA), which recognises the 3-letter codon of mRNA and attaches the corresponding amino acid to the growing protein chain.

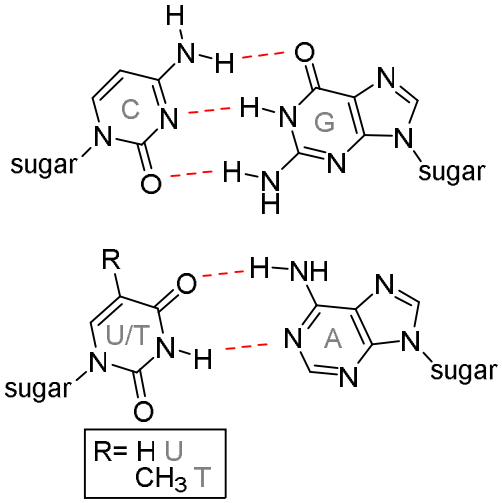


Figure 1.3: Watson-Crick base pairing. Hydrogen bonds are highlighted in red. Sugar represents the ribose/deoxyribose sugar phosphate backbone.

The amino acid pool is ubiquitous (20 naturally occurring amino acids), as is the genetic code (*vide supra*, *Figure 1.1*). The genetic code is degenerate, with most amino acids being coded for by two or more codons (see "family boxes" in *Figure 1.1*, *vide supra*), and only two amino acids, tryptophan and methionine, are coded for by only one codon. The relationship between DNA, RNA and proteins is indivisible, as DNA and RNA are needed for proteins to be formed, and proteins are required for DNA to be replicated and transcribed and for RNA to be translated (through DNA polymerases, RNA polymerases and aminoacyl-tRNA synthetases, to name a few).

After having seemingly deciphered the laws governing molecular biology, namely the central dogma and the genetic code (*vide supra*), it was clear that understanding how the genetic code came about was essential to understanding the origin of modern biology. There are several theories that have been put forward to explain why the genetic code is the way it is. The "frozen accident

theory", as described by Crick, states that "the code is universal because at the present time any change would be lethal, or at least very strongly selected against"[23]. While this might be the case, this theory doesn't account for the (small) variations in the genetic code[i] and the "specific, ordered features of the code"[24]. One other theory postulated by Crick is known as the "stereochemical theory", where the code developed out of inherent stereochemical interactions. In a sister-paper, Orgel went as far as to suggest that perhaps "trinucleotides have a selective affinity for the amino acid coded by their complementary trinucleotide."[25]. The stereochemical theory can be improved by omitting amino acids that have a lengthy biosynthesis, which were deemed to be later additions to biology (such as lysine, tyrosine, phenylalanine, histidine, tryptophan and methionine), thereby simplifying the genetic code to "family boxes" that are independent from the third base of the codon[24] (*Figure 1.4*).

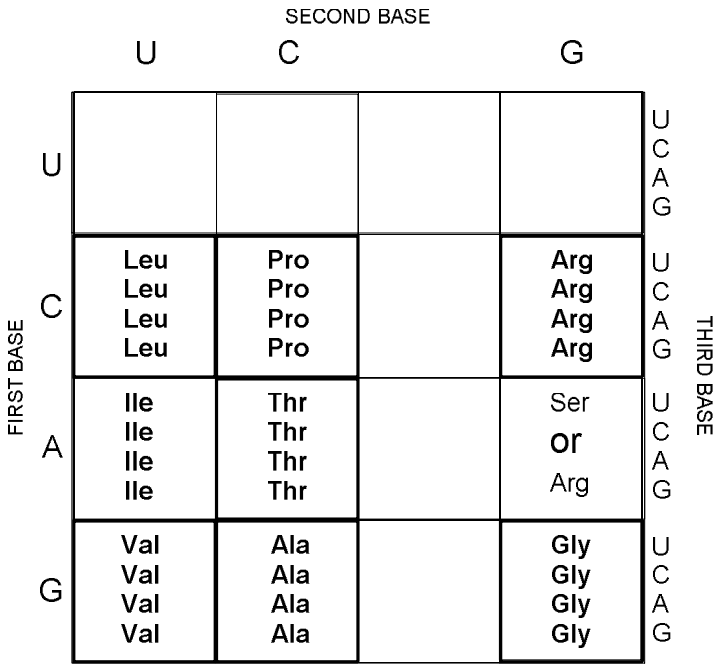


Figure 1.4: Simplified "early" genetic code as postulated by Borsenberger et al.[24]*. XAZ and UYZ codons are removed and reassignement of codons AUG to Ile and AGZ to Ser/Arg. Amino acids are represented by their standard three-letter codes. Simplification of the code forming mainly "family boxes", shown in bold.*

[i] Some small differences are found, mostly in mitochondrial DNA, usually involving a stop codon coding for another amino acid.

Other theories concerning the evolution of the genetic code have been put forward such as the "adaptation theory", which suggests that the extant genetic code was "assigned" to minimise mutations (or mistranslations)[26], or the "co-evolution" or "historical theory" where the code evolved over time when amino acids became available by new biosynthetic routes[27]. Another question that arises from the observed link between the RNA codons and the amino acids coded is "Why is there such a strong relationship between RNA and amino acids?". Proposed explanations include that aminoacylation protected RNA, helped initiate transcription of that RNA[28], or even enhanced the catalytic activity of RNA[29]. These theories suggest that perhaps the genetic code evolved out of this relationship, and therefore was a natural product of the interaction of RNA and amino acids. One thing that all these theories have in common is that they assume extant life evolved from a single population of organisms, known as the Last Universal Common Ancestor (LUCA)[23,30,31]. Indeed, this universally conserved conversation between the nucleotides and amino acids through the genetic code points to a unified origin of life[32–37].

A defining principle of scientific research is that the more we discover on a subject, the more questions we uncover: Why RNA? Why DNA? Why proteins? Why *this* genetic code? When did life start? How did life start? The list goes on. It is most probable that there was no defining point at which inanimate matter became life, but rather, as Sutherland put it, a series of steps in "increase of aliveness" of a system[38]. The goal in the search for the origin of life is to uncover some of those steps and the mechanisms behind them.

1.1.2. Geochemical approach

When considering the origin of life on an abiotic Earth it is important to keep in mind the geochemical context in which it may have arisen. Analysis of the geological record suggests life formed quickly (*Figure 1.5*). Indeed, there is evidence pointing towards microbial life being present about 3.6 billion years ago[34,39] and (debated) sedimentary and isotopic evidence for life 3.8 billion years

ago[40]. The proposed late heavy bombardment of the Earth (and Moon) suggests that extreme energy impacts would have rendered the surface sterile until around 4 billion years ago[41,42].

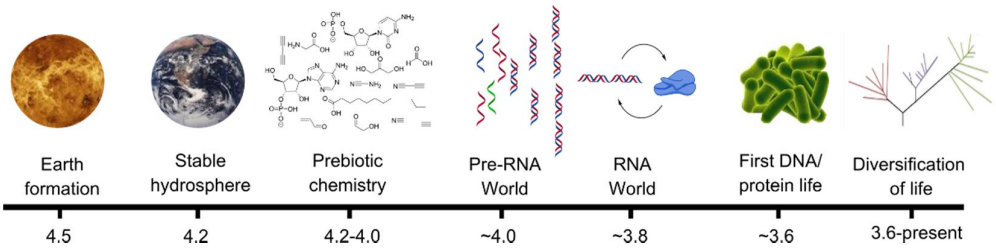


Figure 1.5: Timeline of events leading to the emergence of life. Dates are approximated, in billions of years before present. Adapted from Joyce[34].

The oldest evidence of life on Earth is of *complex* microbial life that is already specialised and established in its environment. This suggests that the origin of life, and the chemical environment leading to self-organisation, membranes, and replication (all of which are observed in even the simplest microorganisms), was established long before. It is with that in mind that people have attempted to explain, through a bottom-up approach, what was so unique about the early Earth's environment and why it was prone to the formation of life[43]. Knowledge of the geochemical history of the Earth is thought to be able to inform the possible chemical reactions that could have taken place and bring essential insights into which direction we should be focussing our efforts into uncovering the origin of life. Consequently, there is much current interest in understanding the sources of compounds that could lead to the development of biomolecules.

One of the possible sources is that of the delivery of compounds from comets and meteorites. While the heavy bombardment (*vide supra*) would have surely delivered a significant amount of materials to the Earth, it is difficult to estimate how important that material was to the inception of life[44]. That being said, a rich mixture of organic molecules have been detected in interstellar space[45] and in carbonaceous meteorites[46–48], with a striking similarity to the mixtures formed through prebiotic chemistry (*vide infra*)[49].

Another possible source is that of the synthesis of compounds from the atmosphere. This route has been explored experimentally through "spark discharge" experiments, designed to mimic the effect of lightning on the atmosphere (*Figure 1.6*). The ground-breaking spark discharge experiment by Miller and Urey in 1953 in a mixture of gases mimicking the (contemporary) early Earth atmospheric model, revolutionised the way the boundary of organic and inorganic chemistry are explored, providing the first simple model for how essential bioorganic molecules might have been formed abiotically[36].

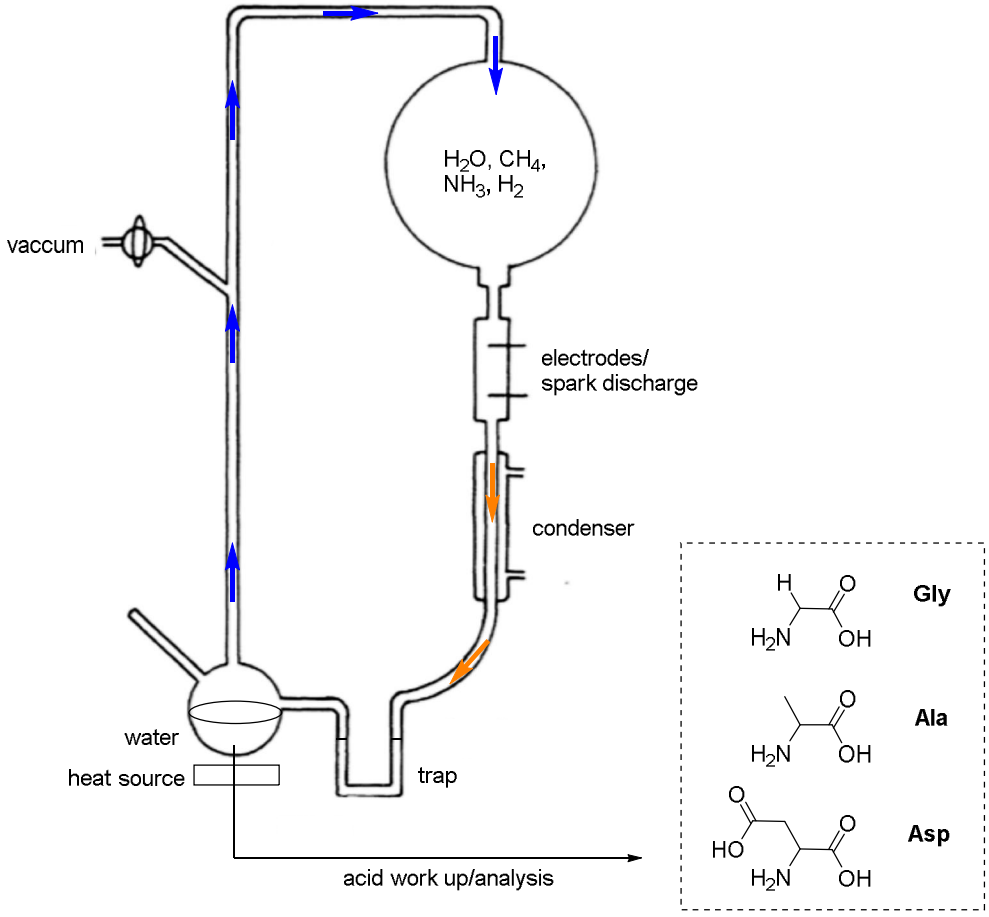


Figure 1.6: Experimentally set-up of the Urey-Miller experiment. Blue arrows represent the circulation of water vapour. Orange arrows represent the condensation of the mixture formed from the spark-discharge experiment. Proteinogenic amino acids formed are shown in the dotted box. Adapted from Miller[50].

It became apparent through this simple experiment that molecules that were previously thought of as too complex to form and could only have been formed by pre-existing life, could self-assemble under laboratory conditions. Abiotically plausible molecules observed, for example in spark discharge and ionisation experiments, have led to the formulation of a repertoire of molecules, including nitriles, aldehydes, amino acids, and other biologically relevant molecules, that may have been present on the early Earth[36,51] (included in this repertoire are some of the 130 molecules that have been detected in space and meteorites, *vide supra*, that could have been delivered to enrich the pool of prebiotic feedstock molecules; *Figure 1.7*). The relevance of the results obtained through these early spark discharge experiments has since been put into question due to the strongly reducing atmospheric model used[52]. It is no longer thought the atmosphere was as reducing as originally thought[53], even though the chemistry described could have occurred to some extent[54].

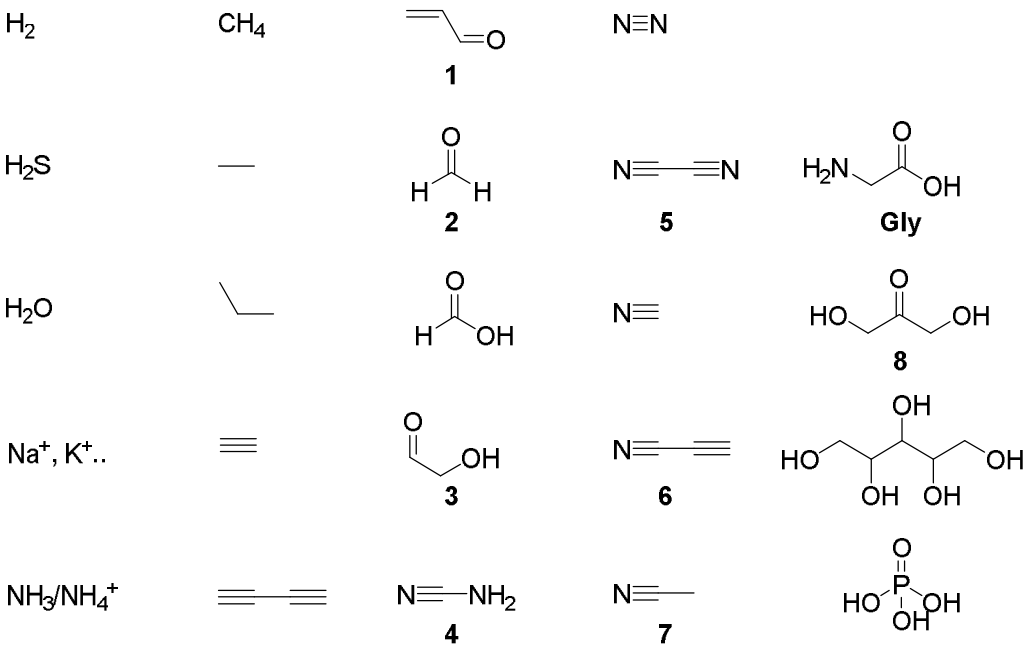


Figure 1.7: Partial repertoire of assumed prebiotically plausible compounds.

Clearly, there is a lot of speculation when it comes to the early atmospheric and geochemical context in which life arose, and efforts have been channelled into

understanding the early Earth conditions, namely through astrogeology and exoplanetary atmospheric composition studies. A number of recent high profile programs such as Kepler and the James Webb Telescope[55] have been planned and developed in order to understand the complex atmospheric conditions of planets outside of our solar system, more and more of which are being discovered every year[56,57]. Hopefully, this will help us improve our understanding of planet formation in general, priceless information which in turn can be applied to models of the early Earth atmosphere, as well as in the search for potential extra-terrestrial life.

On the other hand, some aspects of the geochemical environment are fairly certain. First, the presence of an ultraviolet light (UV) source is undisputed, which we still have present today in the form of our sun. There is evidence that early microorganisms were anaerobic[58,59], suggesting that molecular oxygen was not present in high enough concentration to be harvested for its oxidative potential. Furthermore, the geological record shows significant oxygen buffering capacity in the early biosphere[60] and thus it is inferred that O_2, and by consequence O_3[61] were not present in the atmosphere during biogenesis. Lack of ozone would have altered the light composition on the planetary surface compared to the contemporary Earth, as ozone absorbs UV radiation (which some organisms, such as eukaryotic phytoplankton, don't tolerate in high levels)[60]. Second, it is thought that liquid water was present at the time of abiogenesis as there is geological isotopic evidence[62], and it is considered highly probable that life originated in an aqueous environment[63]. In contrast with the assumption that the early Earth had a strong reducing atmosphere, scientific opinion has recently moved towards a theory of a milder environment with moderate temperatures, low ionic strength and near-neutral pH[64]. Such conditions can be imagined in pools of water, that through cyclic filling and evaporation may have been ideal places for complex chemistry and oligomerisation[65] (*vide infra*, *Sections 1.2.3, 1.2.4 and 1.2.6*).

Experimental design from specific geochemical conditions and theories can lead to interesting results, such as those obtained by Miller[36] (*Figure 1.6*, *vide supra*).

However, inconsistency in theories regarding the atmosphere and environment of the early Earth combined with a restriction on chemical composition has led many scientists to be reluctant to narrow their search based upon the current models for environments that *may* have existed at the time of life's formation.

> *"Starting from geochemistry, planetary science suggests that the early Earth could have offered a wide range of environments and conditions. A huge amount of chemistry is potentially possible in a submarine vent, or a drying lagoon, or an impact crater, or a reduced atmosphere subject to lightning, or whatever other scenario one can imagine. However, there is insufficient constraint from geochemistry* per se *to settle on one particular scenario, and thence to systematically explore its chemistry with a view to uncovering intrinsically favoured syntheses of biomolecules, or their precursors, from simple feedstock molecules."* - Sutherland[38]

Nevertheless, it is apparent that the early Earth environment would have been rich in organic molecules and prone to further assembly of more complex prebiotic molecules. Many questions remain: how did these molecules assemble to form a living system? Will the chemistry point us in the direction of particular geological environments? Will we ever know where life started?

1.1.3. Biological approach

Current and past life gives us a few hints towards early biology, and a lot of researchers have considered the problem of biogenesis from this top-down approach: looking at biology in order to glean some information that could be revealing. As discussed previously, the uniformity of molecular biology suggests all extant life is descendant from a last universal common ancestor (LUCA). Analysis of ribosomal RNA genes (16S rRNA) has helped reconstruct phylogenies of early life[66] (*Figure 1.8*). Protein sequence analysis has also helped

determined that LUCA lives 3.2-3.8 billion years ago and was comparable to modern bacterium in terms of complexity[30].

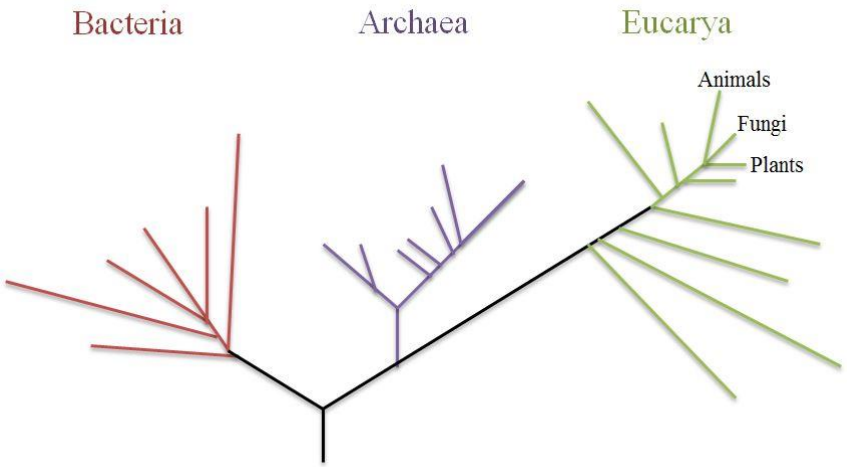


Figure 1.8: Simplified universal phylogenetic tree of terran life, adapted from Woese[67].

The techniques of paleogenetics can be also used to resuscitate extinct genomes and early proteins. One study by Gaucher *et al*. successfully resurrected a 3 billion year old protein[68]. The information gathered from property analysis of this protein, namely that the temperature optima for elongation factors is around 60°C, suggests the organism was a thermophile[68]. This has been interpreted as giving credence to theories of a deep-sea hydrothermal vent origin of life (*vide infra, Section 1.1.3.1*). However, the phylogenetic evidence could simply be a remnant of a bottle-neck episode where surface sterilisation occurred, leaving the deep-sea organisms to colonise the earth[69]. These paleogenetics studies have revealed interesting information, however this approach is seriously limited by prolific horizontal gene transfer (transfer of genes by means other than traditional reproduction) in the (early) microorganism biosphere, which causes ambiguity in the assignment of ancient phylogeny[70].

Extant and extinct biology corroborates information regarding essential characteristics of life. The minimal cell approach to studying biogenesis looks at

those essential (universal) aspects, namely hereditary information (e.g. DNA), a source of building blocks (e.g. nucleic acids and amino acids), compartmentalisation (e.g. membranes), chemical energy (e.g. metabolism) and catalysts (e.g. proteins and enzymes)[71,72]. Most research concentrates on one aspect as it is usually easier to categorise them for research purposes.

1.1.3.1. Metabolism

The metabolism-first approach to life is based on an autotrophic theory of evolution where the drive behind the origin of life was energy, rather than information (*vide infra*, *Section 1.1.3.3*). Some supporting information towards the idea of an autotrophic origin of life is the fact that the selection observed in the glycolysis pathway can be obtained spontaneously simply by heating each substrate in an "Archaen ocean" model[73]. Furthermore, some argue that the lack of obvious nucleotides involvement in central metabolic networks hints that life was not dependent on RNA/DNA at the start[74].

The metabolism-first approach has gained much interest since Wächtershäuser suggested an autotrophic theory of evolution wherein the energy necessary for the building and maintenance of life, which in itself is an entropically unfavourable process, is stored in reducing iron species[75]. This theory is retrodictively based on the information gathered from some extant biochemical processes such as the citric acid cycle, but differs in that it replaces the source of the reducing power NADH by pyrite FeS_2, and thioesters by thioacids[75]. Extant chemoautotrophic bacteria utilising transition metal sulfides in their metabolism have been interpreted as evidence that iron-sulfur metabolism could have been present in the first organisms[76,77]. Wächtershäuser proposed the formation of pyrite from ferrous sulfide (FeS) and hydrogen sulfide (H_2S), and postulated that "the geochemical setting of the origin of the archaic precursor cycle is an anaerobic aqueous environment, rich in H_2S (or HS^-) and in contact with heavy metal sulfides and pyrite" such as one present in deep-sea hydrothermal vents[75]. The possible reducing power of H_2S and FeS_2 has been demonstrated experimentally in the reduction of thiols and acetylene by H_2S/FeS and successful nitrogen fixation

from the reduction of N_2 to ammonia, NH_3[78,79]. Autocatalytic metabolic cycles and primitive metabolisms have also been linked experimentally with chiral amplification[80] and amino acid production[81,82] (*vide infra*, *Section 1.1.3.3*).

The theory of a deep-sea hydrothermal vent origin of life is still actively discussed today[83,84]. Evidence potentially pointing in the direction of a hydrothermal vent origin include the observation that the oldest known microbes include a vast number of thermophilic species[85]. Furthermore, energy-rich pH gradients of hydrothermal environments could have been an ideal environment for establishing proton gradients, that some view as central to metabolism and life[84]. These proton gradients in geothermal fields[86,87] (theoretically of up to 4 pH units between the alkaline hydrothermal vent pore system and the acidic ocean[88]) have been used as energy sources in theoretical models of an autotrophic origin of life, driving the energy-consuming metabolic processes and eventually leading to the onset of life[89]. Furthermore, the inorganic ion distribution is found to be similar in the effluent of geothermal fields to that found in cells[84,90]. Other arguments in favour of a geothermal origin of life include the fact that it is independent of surface environment and that there is a (relative) abundance of phosphate[90] (*vide infra*, *Section 1.2.5*).

However, these theories have received a lot of criticism. Firstly, the vastness of the oceans in which such processes would take place leads to a discrepancy which has been termed the "concentration problem of the origin of life"[91]. A theoretical solution has been proposed by Baaske *et al.* in which elongated hydrothermal pores would lead to concentration of large molecules by means of thermal convection and diffusion, where the added benefit of thermal oscillation could lead to an increase in oligomerisation rates of nucleotides[91]. This model has been applied to concentrate dilute solutions of biologically relevant molecules such as fatty acids (essential in membrane composition) and nucleotides. The study by Budin *et al.* showed self-assembly of vesicles encapsulating nucleotide oligomers, confirming the possibility of concentrating prebiotic feedstock molecules in this dilute environment[92]. Nevertheless, Lazcano maintains that these environments

would be limited in their abundance of organic molecules as it is much less efficient in producing biomolecules than spark discharge experiments[93]. Furthermore, the proton gradient/hydrothermal vent hypothesis not only lacks experimental backing, but also suffers from severe flaws when it comes to the actual potential of energy available from such pH differences[94]. Additionally, the observed thickness of the inorganic barriers that separate the vents from the ocean are of the order of 200 times the thickness of biological membranes[86,87], making the potential proton gradient difficult to access[94].

1.1.3.2. Membranes

"Molecules that stay together evolve together"[63] perfectly sums up the membrane-first approach to life. It is undeniable that life on Earth would not have evolved to the point it is now without compartmentalisation of prebiotic molecules or primitive biopolymers, which allow concentration of compounds as well as allowing that compartment, or cell, to have a higher chemical energy potential and lower entropy than its surrounding[35,95]. Furthermore, Darwinian evolution is greatly facilitated by compartmentalisation as it can co-localise genotype and phenotype[24,35,96]. Separating the compartment from the exterior could also procure protection from "parasitic" molecules[97].

If we consider primitive biomolecules, such as RNA or proteins, which are utilising small molecule metabolites and building blocks, and in turn producing others necessary to the function/fitness of this primitive system, it seems obvious that these small molecules would need to be in close enough proximity for them to be used efficiently by the system. Retaining these small molecules long enough for their use by other biopolymers or macromolecules would thus be essential to the efficient functioning of the system. The same applies to the biomolecules themselves, as they can only interact if they are in close proximity. However, the nature of the compartmentalisation necessary for chemical evolution is debatable. A semi-permeable membrane would be ideal in allowing small molecules and metabolites from the surroundings to participate in the system's molecular feedstock, and allowing waste to be excreted, but would also allow for substantial

loss of material. On the other hand, an impermeable membrane would allow full retention of these small molecules but could only be advantageous if the system is self-sustaining in its metabolite feedstock, which is not possible over an extended period of time. To account for this apparent paradox, Orgel suggested a time-line for membrane formation with an evolving mixed character[63], where a permeable membrane would originally be more profitable for chemical evolution and impermeability would increasingly become a 'desirable' attribute. One could also imagine a changing permeability such as the mechanical stress-induced liposome permeabilisation observed by Natsume and Yoshimoto[98].

Experimental studies have been conducted towards the prebiotic plausibility of membrane formation in an aqueous environment. These experiments show that there are several prebiotically plausible ways to form vesicles and lipid bilayers, both of which are ubiquitous to biological membranes. There is also evidence for the synthesis of long-chain fatty acids in prebiotic conditions[99] and their spontaneous formation of vesicular structures[100,101]. Amphiphilic compounds formed experimentally in plausible interstellar conditions (through the irradiation of interstellar ice), and those detected in the Murchison meteorite[102], can lead to spontaneous vesicle formation[102,103].

Some minerals, such as montmorillonite, can help template closed vesicle formation[104]. Hanczyc *et al.* remarked on the advantage of montmorillonite-assisted vesicle formation as particles of the clay mineral trapped during vesicle formation can import their (potential) catalytic activity into the enclosed system[104]. It is of note that effective mineral-catalysed oligomerisation of activated nucleotides has been demonstrated (a point to be returned to later, *Section 1.2.6*). Furthermore, compartmentalisation has been shown to improve RNA-directed catalysis by "molecular crowding"[105]. Some processes have previously been seen as incompatible with membranes. For example, a high concentration of divalent cations such as Mg^{2+}, which are essential to RNA copying[106], leads to vesicle degradation by precipitating the fatty acid present in the membrane[107]. Fortunately, a solution was found by Adamala and Szostak in

the form of small molecules Mg^{2+} chelators, such as citrate, that protect the vesicle from degradation while allowing RNA oligomerisation to occur efficiently[107].

The advantages of compartmentalisation are numerous, however it is difficult to imagine how life could have started from membranes alone[24,108]. The need for membranes suggests a co-evolution with one or more systems[35]. On that note, a link between sugar selectivity and membrane permeability has been studied by Sacerdote and Szostak[109]. They have demonstrated that ribose (**9**) has a kinetic advantage when it comes to assimilation in proto-cells with a 3-10 fold increase in diffusion rate compared to other aldopentose sugars, which could have been a "factor that facilitated the emergence of the RNA world"[109] (*vide infra*, *Section 1.2*). On the other hand, membranes could also have been preceded by one of these other systems, which might explain why the membrane lipids of different "branches" of the tree of life are different (while the catalytic and informational biomolecules are ubiquitious)[110].

The apparent need for *bilayer* encapsulation has been debated. Orgel, for example, believed the same advantage could be gained environmentally, by co-localisation[63]. There is a wealth of theoretical studies[81], as well as experimental evidence[83,102], that suggests rocks and minerals present on the early Earth could have increased oligomerisation rates and local concentration of important biomolecules, such as RNA and proteins, and thus surpasses the need for the presence of membranes in the early days of chemical evolution. Other suggestions include encapsulation by something simpler than phospholipids that might have been more available[111,112]. Regardless of when membranes came into play in the evolution of life, it is undeniable that they were necessary at some point in its conception.

1.1.3.3. Biopolymers

The biopolymer-first approach to biogenesis has been a fundamental part of the research into the origins of life due to the obvious centrality of DNA, RNA and

proteins in the cell's function. The heterotrophic theory of evolution, where biopolymers are used to build life, encompasses many sub-theories ranging from a genetic central theory (where hereditary information was the root of life) to a catalytic centred one (where catalytic activity such as the one displayed by proteins and enzymes was the first essential feature of life to arise), or as Orgel put it:

> *"If we accept that the biochemistry of the earliest organisms was recognizably related to that which we now know, there are only two obvious alternatives to the present genetic system which need to be considered, namely : (1) life based on proteins in the absence of nucleic acids, (2) life based on nucleic acids in the absence of proteins."*- Orgel[25]

A third-theory, based on an alternative biopolymer pre-dating proteins and RNA/DNA will be discussed at a later point (*Section 1.3.1*). The importance and shortcomings of RNA based theories to the origin of life will be considered in depth in *Section 1.2*, but first other proposals will be discussed.

One school of thought advocates that life began with self-replicating proteins. Oparin went so far as to say, "the origin of life must have coincided with the origin of proteins if not actually preceded by them"[59]. The prebiotic plausibility of many amino acids has been verified through spark discharge experiments and analysis of chondritic meteorites[36,48] (*vide supra*). The indisputable prebiotic accessibility to amino acids, coupled with their chemical simplicity makes this theory highly attractive (especially in light of difficulties in early prebiotic nucleotide synthesis, a point to be returned to, *Sections 1.2.1-1.2.5*). The oligomerisation of amino acids into peptides is conceptually easy (condensation between a carboxylic acid and amine), however it is rendered difficult in a prebiotic context due to the pK_a mismatch between the two moieties: carboxylic acids requires low pH to be electrophilic ($pK_a \sim 4$) whereas amines needs high pH to be nucleophilic ($pK_a \sim 10.5$; *Figure 1.9*).

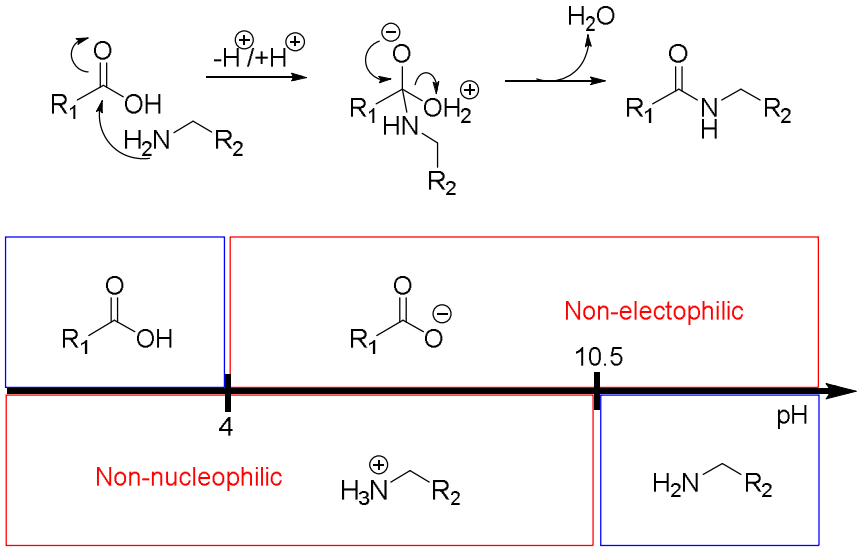


Figure 1.9: Mechanism of peptide bond formation (top) and diagram of predominant protonation states of carboxylic acids and amines at different pHs (bottom).

Nevertheless, proteins are essential to the function of cells as a major structural and catalytic component. Protein function is both ubiquitous and essential, for example in signalling, DNA repair and replication, membrane maintenance and is central in metabolism[22]. Proteins are absolutely indispensable to modern life, but the question remains as to whether they were needed in its emergence. Great advances have been made with regards to peptides participating in biogenesis. The Ghadiri group demonstrated peptide copying through a molecular level symbiosis in the form of a coiled-coil peptide hypercycle, where the production of one peptide catalyses the production of another, as well as feeding back into its own production (non-linear autocatalysis)[113]. Additionally, small peptides have also been shown to have catalytic activity[114] and self-replicating abilities[115], albeit in a very sequence specific manner. However, the use of proteins and peptides as the basis for the origin of life is debated as their disposition to form secondary and tertiary structures makes accessing the amino acid sequence difficult and results in a loss of the templating activity necessary to replication of information[31].

There is a school of thought that proposed there must be a reason that life 'chose' two (constitutionally) highly distinct biopolymers, or in other words, that there

may have been an evolutionary advantage to having nucleic acids and amino acids (and other important metabolites) interact with each other. As such, the relationship between proteins and DNA/RNA has been explored in some depth. Extant biology provides a number of clues to the deep-seated relationship between peptides and nucleic acid. One good example is that of RNA polymerases, the enzymes necessary for transcription of DNA into RNA. At the active site of these enzymes are metal ions[116] coordinated by aspartic acid residues - a common amino acid formed in spark discharge experiments[36,117] and detected in meteorites[118] - contained within a highly conserved amino acid sequence[119]. Szostak *et al.* have postulated that random sequences of amino acids, formed on the early Earth, could have led to catalytic sequences, which would be able to contribute to the activity of primitive catalytic molecules such as ribozymes (*vide infra*), by coordinating and stabilising divalent metal ions[120]. Indeed, metal ions can also be useful in enhancing, or are often essential to, catalytic activity of ribozymes[121]. Another biological link between nucleic acids and amino acids is tRNA. Indeed, amino acyl-tRNA and amino acyl adenolates both utilise direct covalent conjugation of amino acids and nucleotides to fulfil their roles in biochemical information transfers[22]. Interestingly, the Schimmel group has shown that in a ribosome-free environment, tRNA mini-helix could catalyse the formation of peptide bonds in the presence of an imidazole (likely acting as a general base) *via* direct interaction of two aminoacylated oligonucleotides[122]. Another interesting aspect of endogenous tRNA is its apparent discrimination against D-amino acids, leading some to suggest that the tRNA aminoacylation may have been the initiator of the homochirality observed in present-day amino acids[123] (*vide infra*). When it comes to the abiotic synthesis of amino acids and nucleotides (or their precursors), there is a lot of overlap of materials used. Indeed, conditions that have been found to produce purine nucleobases, through HCN oligomerisation (*vide infra*, *Section 1.2.3*), also leads to the synthesis of some of the 20 canonical amino acids[124] (a point to be returned to, *Section 2*). Photochemistry of HCN can also lead to synthesis of amino acids, nucleotides and membrane precursors[37] (a point to be returned to later, *Section 1.3.2*). This facile

co-synthesis of the respective building blocks of proteins and nucleotides and their constitutional relationship gives weight to a heterotrophic origin of life[38].

Why did life "choose" one enantiomer? Homochirality, like the one present in the two main biopolymer building blocks of life on Earth, L-amino acids and D-(deoxy)ribonucleic acids, has always been a big question in the origin of life. This is an essential problem to solve as template-directed polymerisation of RNA is stopped by the incorporation of the opposite enantiomer, meaning that racemic mixtures lead to dead-end oligomers[125,126]. Interestingly, some meteorites show a significant enantiomeric excess (*ee*) in favour of the canonical L-amino acids, such as the Murchison meteorite where L-isovaline is observed in up to L_{ee}=18.5%[127]. Similar L_{ee} are observed in the Murray meteorite (isovaline L_{ee} up to 6%)[128] and in the Orgueil meteorite (isovaline L_{ee} up to 15.2%)[127]. Perhaps life didn't 'choose' an enantiomer, but was instead imposed one?

Nevertheless, the question remains of how to go from a slight *ee* to selectively obtaining an enantiomerically pure set of building blocks and biopolymers. Non-prebiotic experiments into "breaking symmetry" and homochirality have been studied, such as those by Soai *et al.*[80] and Viedma[129]. The positive feedback autocatalytic reactions demonstrated by Soai *et al.* where reaction of 5-pyrimidyl alkanol with low *ee* enrichment is treated with diisopropyl zinc to yield 5-pyrimidyl alkanol with high *ee* enrichment, showed experimental validation to the theory of *ee* amplification. Crystal grinding studies by Viedma showed that achiral sodium perchlorate spontaneously crystallises into a chiral lattice, and through grinding and recycling eventually leads to complete chiral purity in the crystal phase[129,130]. This crystal grinding technique was later applied to prebiotic chemistry by Blackmond and colleagues, showing successful chiral amplification when starting with a low L_{ee} of aspartic acid[131]. The solid-liquid phase behaviour of amino acids has also been invoked to enriching the L_{ee} of amino acids. When starting with a low *ee* (1%) of phenylalanine, Breslow and Levine observed selective crystallisation of racemic crystals (through slow evaporation), leading to 90% *ee* in the solution after only two rounds of evaporation[132]. Similar results

were obtained by Klussmann *et al.* with other proteinogenic amino acids, where near-racemic mixtures lead to high *ee* in the solution phase[133]. The use of the *ee* amplification of amino acids has also been tentatively applied to carbohydrate synthesis through the asymmetrical amino acid-catalysis of hexose sugars[11]. These studies show that through prebiotically plausible means or not, breaking symmetry is not an impossible feat.

The interest in bridging the gap between the geochemical approach, metabolism-based and template-based approaches to the origin of life shows that an integrated theory needs to be found. This holistic approach to biogenesis attempts to think of the overall system and see the benefits of combining all fields of research. This approach, known as systems chemistry, studies complex mixtures in order to better understand the way the complex chemical environment can affect the outcome of the reaction[135,136] (*vide infra*, *Section 1.3.2*). Before we delve into the systems chemistry world, first we must explore another subject only alluded to so far: RNA.

1.2. The RNA world theory and its obstacles

RNA is an essential molecule to all life on Earth. However, it was only recently discovered to what extent life *relies* upon RNA. Half a century ago, RNA was thought of only as the link between DNA and proteins, helping convey the information from one to the building of the latter, through mRNA and tRNA[22] (*vide supra, Section 1.1.1*). However, the discovery that RNA also has an indispensable catalytic role in cells[137], led to a revival in the school of thought that RNA was the first biopolymer of life: the 'RNA World' theory (coined by Gilbert in 1986[96]).

> *"It may be claimed, without too much exaggeration, that the problem of the origin of life is the problem of the origin of the RNA World, and that everything that followed is in the domain of natural selection"*- Orgel[63]

Even half a century before catalytic RNA was discovered, it was considered a possibility that RNA might have been essential to the origin of life[63,138]. Crick and Orgel both proposed that there might be a biopolymer that could have both genotypic and catalytic properties[23,25], thereby providing a solution to the paradox of "Which came first DNA or proteins?". RNA seemed a likely candidate, based upon several clues that it was present very early on in evolutionary development, clues that Francis Crick described as giving him an overwhelming feeling that RNA was part of the primitive machinery of life[23]. Its centrality and ubiquity in biology hints at a possible involvement in the origin of life. Indeed, like DNA, RNA can store and replicate information; RNA primes DNA to initiate the latter's replication[22]; RNA is found as an essential component of many of the main cellular machinery such as the telomerase ribonucleoprotein[139], the ribosome[137] and the spliceosome[22]; RNA is essential in the manufacturing of proteins as it acts as a template for protein assembly in the form of mRNA[22]; RNA acts like a link between the template and the building of the amino acid sequence of proteins, in the process of protein translation, through tRNA[22]. Crick even described tRNA as "Nature's attempt to make RNA do the job of a protein"[23]. All of these roles combined show that RNA is much more versatile than DNA. Interestingly, deoxyribonucleotides are made from ribonucleotides in cells, suggesting RNA is an older structure than DNA[140,141]. Furthermore, the ribosome (the ribonucleoprotein that translates mRNA into proteins, *vide infra*) is homologous across the domains of life, as opposed to DNA polymerases (the enzymes that build DNA from deoxynucleotides) that are non-homologous across the tree of life[22], pointing into the direction of a later emergence of DNA[38].

Another point of interest is that ribonucleotides are also present in some essential co-factors such as ATP, FAD, Acetyl-CoA and NAD^+/NADH[142] (*Figure 1.10*), where the role of ribonucleotide moiety is unclear. It has been suggested that nucleic acid moieties are molecular fossils of an earlier stage in evolution where they might have been indispensable to metabolism[95,143–145], again suggesting ribonucleotides and RNA could pre-date proteins.

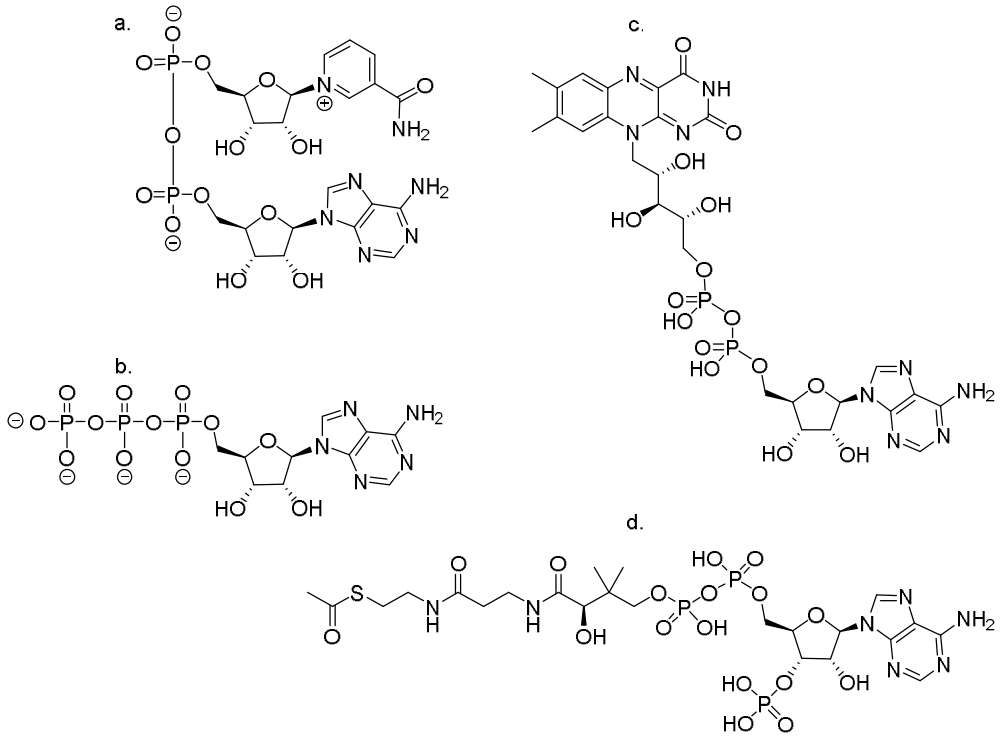


Figure 1.10: Cofactors containing ribonucleotides. a. Nicotinamide adenine dinucleotide (NAD); b. Adenosine triphosphate (ATP); c. Flavin adenine dinucleotide (FAD); d. Acetyl coenzyme A (acetyl-CoA).

Additional to its essential function in contemporary cells, RNA as a biopolymer has characteristics that make it a strong candidate for an early-life biopolymer as RNA templates are able to direct non-enzymatic synthesis of complimentary strands[146].

Most importantly, as Neveu *et al.* said, RNA is "quite unusual among polymeric systems in its ability to strike a balance between contradicting needs of catalysis and genetics"[142]. Indeed, advances in the last few decades have shown that RNA possesses catalytic abilities through a vast repertoire of reactions, such as self-splicing[147,148], RNA cleavage[149], ligation[150,151], phosphorylation[152] and a variety of other catalytic activities[153].

What really cemented RNA's dual functionality as both a genetic and phenotypic biopolymer was the unexpected discovery that "the ribosome is a ribozyme"[154]. While the ribosome was known to be a ribonucleoprotein, the RNA moiety was

thought to have more of a scaffolding role. However, it was revealed that RNA is at the heart of the catalytic activity[137] of ribosomes as no protein residue come within 18 Å of the peptidyl transfer site[155], and RNA alone can catalyse the peptide bond formation[156]. This functional duality of RNA endows it with the capability of Darwinian evolution at the molecular level, which has been demonstrated in the field of *in vitro* evolution[157–162], and in the evolution of viral RNA[163].

The evolutionary capabilities of RNA and its potential for acting as both genotype and phenotype has led to RNA being referred to as "the molecular biologist's dream"[164]. In the "RNA World" theory it is imagined that RNA was produced abiotically, followed by self-replication, until more complex systems were evolved such as a ribosome and/or protein synthesis.

> *"The invention of protein synthesis, instructed and catalysed by RNA, was the crowning achievement of the RNA world, but also began its demise."*- Joyce[34]

As well as solving the paradox of which came first, information or catalysis, the "RNA world" theory also simplifies the problem of the origin of life by avoiding the "double improbability" of two complex biopolymers evolving simultaneously in a complex co-dependent relationship[142]. This is not to say that RNA functioned solely on its own, as there is still scope within the theory for interactions with other biomolecules (*vide infra*, *Section 1.3.2*).

Despite the body of evidence that biological dependence on RNA is much more deeply rooted than originally thought, explaining the prebiotic synthesis of such a complex molecule as RNA is another matter. Until recently, research conducted on the plausibility of making RNA from prebiotic precursors had focused on the obvious (retrosynthetic) approach separating the base, sugar, and phosphate of the mononucleotide (*Figure 1.11*).

"In terms of their appearance to the human eye, ribonucleotides undoubtedly consist of these three building blocks" - Sutherland[165]

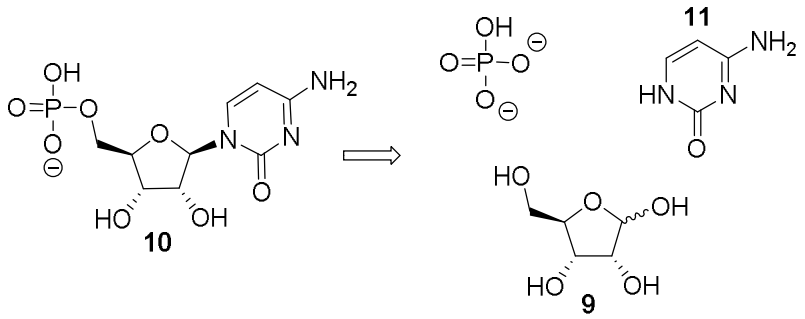


*Figure 1.11: Traditional retrosynthetic analysis of ribonucleotides as demonstrated on ribocytidine (**10**) into the three building blocks ribose (**9**), cytosine (**11**) and inorganic phosphate.*

As we will see in following sections, it wasn't long before the obstacles surrounding the emergence of RNA in an abiotic environment arose, leading some to claim that RNA is not prebiotically plausible and calling the event leading to its nascency "extremely improbable"[166], referring to it as the "prebiotic chemist's nightmare"[167]. However, it is noteworthy that regardless of whether RNA was the first genetic polymer or not, it plays a key role in modern biology and therefore understanding how it arose from an abiotic world is essential.

1.2.1. Constraints of prebiotic chemistry

There are many stereochemical constraints to take into account when looking at the prebiotic synthesis of RNA (furanose not pyranose, ribose not arabinose, lyxose or xylose, 3′,5′ phosphodiester bond not 2′,5′, etc. *Figure 1.12*). But what does prebiotic synthesis entail? Prebiotic chemistry is the field of study which aims to unravel the chemical pathways that led to the emergence of life. In the case of RNA, the aim is to explain how the nucleotides could have emerged from the pool of molecules deemed prebiotically available (*Section 1.1.2*) to form RNA selectively, while taking into account the selectivity issues mentioned above.

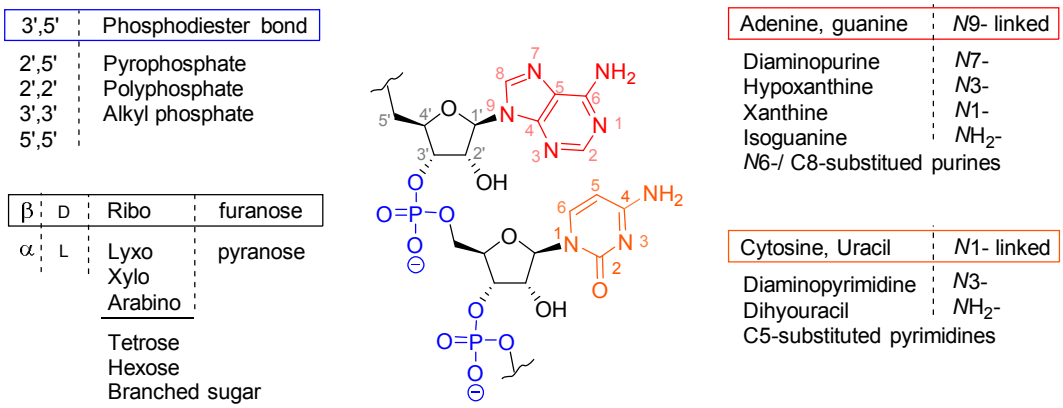


Figure 1.12: Stereochemical constraints on the synthesis of RNA, adapted from Joyce[34].

As Shapiro perfectly put it: "Unfortunately, neither chemists nor laboratories were present on the early Earth to produce RNA"[168]. That is to say that some constraints are imposed on the prebiotic chemists as to whether a certain synthesis could be plausible. Guidelines were devised by Orgel and Lohrmann concerning prebiotic chemical research, which are adapted below[169]:

1- All principle organic reagents must be accessible on the early Earth (*Section 1.1.2*)
2- Water must be used as the primary solvent, and the reaction must be carried out at moderate pH
3- Solid state reactions must occur without excessive drying
4- Physical conditions (temperature, pressure) must be available on the Earth today (no more than ~90°C)

Orgel later added to these that the starting materials should have been accessible in "adequate amounts at the site of synthesis" and that the yield of the desired product in a reaction "must be 'significant', at least in the view of the proposers of the synthesis"[63]. It is clear from these "rules" that there is some leeway concerning what is deemed "available", "significant" and "prebiotically plausible", however the use of certain compounds and conditions need to be

reasoned and defendable if needed. It is also clear from this list that prebiotic chemistry is much more restricted than conventional chemical synthesis, with the use of only one solvent, what would be considered "mild" conditions (usually room temperature (RT) and neutral pH), and no complex protecting group strategies.

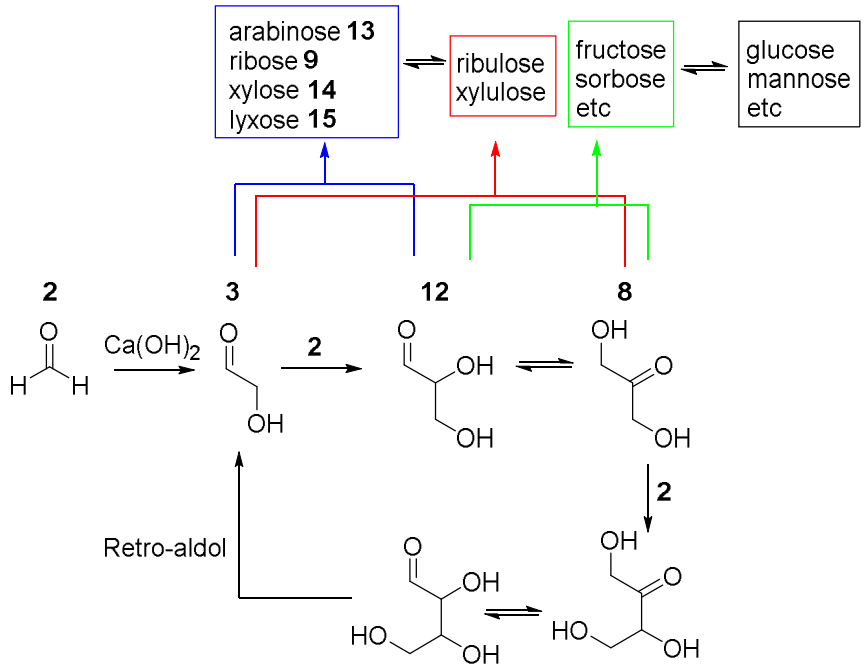
Obviously, it is impossible to accurately recreate the exact conditions and reagents available on the early Earth. As chemists, we are also restricted by both the time-frame of the work carried out, as well as the analytical options at our disposal. These limitations have led to criticisms concerning to what extent the results obtained in a laboratory can be "projected" to the origin of life, best summed up in this quote by Shapiro:

> *"The analogy that comes to mind is that of a golfer, who having played a golf ball through an 18-hole course, then assumed that the ball could also play itself around the course in his absence. He had demonstrated the possibility of the event; it was only necessary to presume that some combination of natural forces (earthquakes, winds, tornadoes and floods, for example) could produce the same result, given enough time. No physical law need be broken for spontaneous RNA formation to happen, but [...] the suggestion implies that the non-living world had an innate desire to generate RNA."* - Shapiro[168]

Nevertheless, through prebiotic chemical analysis of inherently favoured chemical pathways, we can perhaps come closer to understanding what environments are "pre-disposed" to the evolution of complexity and to explaining how life *might* have originated.

1.2.2. Ribose

One of the three components resulting from a traditional retrosynthesis of RNA (*vide supra*) is the sugar moiety: ribose (**9**). Prebiotic synthesis of **9** has been demonstrated since 1861 in a reaction known as the formose reaction[170]. In this reaction, discovered by Butlerov, it is proposed that complex sugars, including **9**, are formed from the initial aldol reaction of glycolaldehyde (**3**) and formaldehyde (**2**) through an autocatalytic cycle, catalysed by calcium hydroxide (*Scheme 1.1*)[171].



Scheme 1.1: Complex racemic product mixture as formed in the initial formose reaction.

Formaldehyde (**2**) is the simplest aldehyde, with only one carbon, and is both an abundant interstellar molecule[172–174] and is readily formed in spark discharge experiments[175–177]. As such, **2** is deemed prebiotically available[178]. However, there are two main problems with the traditional formose reaction. Firstly, it has been suggested that the formose reaction occurs thanks to trace impurities of "biologically derived sugars", such as **3**, in the formaldehyde used, that act as "initiators" of the reaction[179]. Without these initiators, **2** will go down a different

pathway through the Cannizaro reaction, forming methanol and formic acid instead[179]. Secondly, and most importantly, ribose (**9**) is only a small proportion of the complex sugar mixture[180], including degradation products (*vide supra*, *Scheme 1.1*), all of which compete with **9** in future reactions towards the synthesis of RNA. Improvements on the formose reaction have therefore been attempted, in order to selectively amplify the proportion of the desired sugar **9** in the mixture, as described below.

Zubay found that replacing the traditional calcium hydroxide catalyst in the formose reaction (in the presence of an initiator: glycolaldehyde (**3**) or dihydroxyacetone (**8**)) with Pb^{2+} ion, in a suspension of magnesium hydroxide, lead to an increase of pentose sugar, with a reported 30% maximum yield[181]. The prebiotic abundance of Pb^{2+} is argued through the presence of lead in numerous minerals[181]. Another way to improve the selectivity of the formose reaction is to start with pre-formed **3** and glyceraldehyde (**12**), the proposed precursors of the pentose sugars in the formose reaction[182]. Both can be formed through aldolisation of formaldehyde (**2**) and the C_2 aldehyde **3**, the latter of which was also shown to be readily formed in spark discharge experiments in a wet CO_2 atmosphere[183].

Ricardo *et al.*'s work is based on the use of borate minerals as catalysts[184]. They report increased formation of the pentose sugars in the aldol reaction of **3** and **12** in the presence of borate minerals. However, even though pentose sugars are stabilised in the presence of borate, the yield and stereoselectivity of their formation is unreported[184]. Zinc-proline has also been studied as a potential catalyst in the same reaction of (equimolar) **3** and **12**, showing an improved selectivity of 20% ribose (**9**; alongside the other pentose sugars and tetrose and hexose sugars)[185,186]. The most *ribo*- selective variant of the formose reaction involves the use of phosphorylated starting material glycolaldehyde-phosphate (**16**). Müller *et al*. first demonstrated that glycolaldehyde-phosphate (**16**), in the presence of 0.5 eq formaldehyde (**2**), leads to the formation of ~25% ribose 2,4-diphosphate (**17**)[187]. These phosphorylated sugars have the advantage of being more stable than the free counterpart and are reminiscent of the fact that sugars

are always used in their phosphorylated form in biology[32]. Free sugars suffer from a short half-life (ribose (**9**): 300 days at 25°C and neutral pH[188]), making them essentially unavailable for prebiotic use and prone to degradation in the conditions of their synthesis as discussed this far. The starting material in this reaction, **16**, can be prebiotically formed from the reaction of **3** with amidotriphosphate (AmTP)[189] (*vide infra*). The disadvantage of this route to **9** is the phosphorylation pattern. Indeed, the sugar obtained is the non-canonical pyranosyl ribose 2,4-diphosphate (**17**), and so far no prebiotic way of no converting pyranosyl **17** to the canonical furanosyl ribose 3,5-diphosphate has been found. Eschenmoser argued for the use of pyranosyl **17** as the sugar phosphate backbone in an alternative biopolymer p-RNA (a point to be returned to later, *Section 1.3.1*).

These results suggest some prebiotically plausible mechanisms allow access to ribose (**9**), although only transiently as it is easily degraded in the same conditions in which it is produced. Another problem raised in the prebiotic synthesis of **9** is the fact that "nitrogenous substances (needed for prebiotic base synthesis) would interfere with the formose reaction by reacting with formaldehyde (**2**), the intermediates, and sugar products in undesirable ways."[190]. This has led to the statement that "neither ribose nor any carbohydrate could possibly have been a prebiotic genetic molecule"[188].

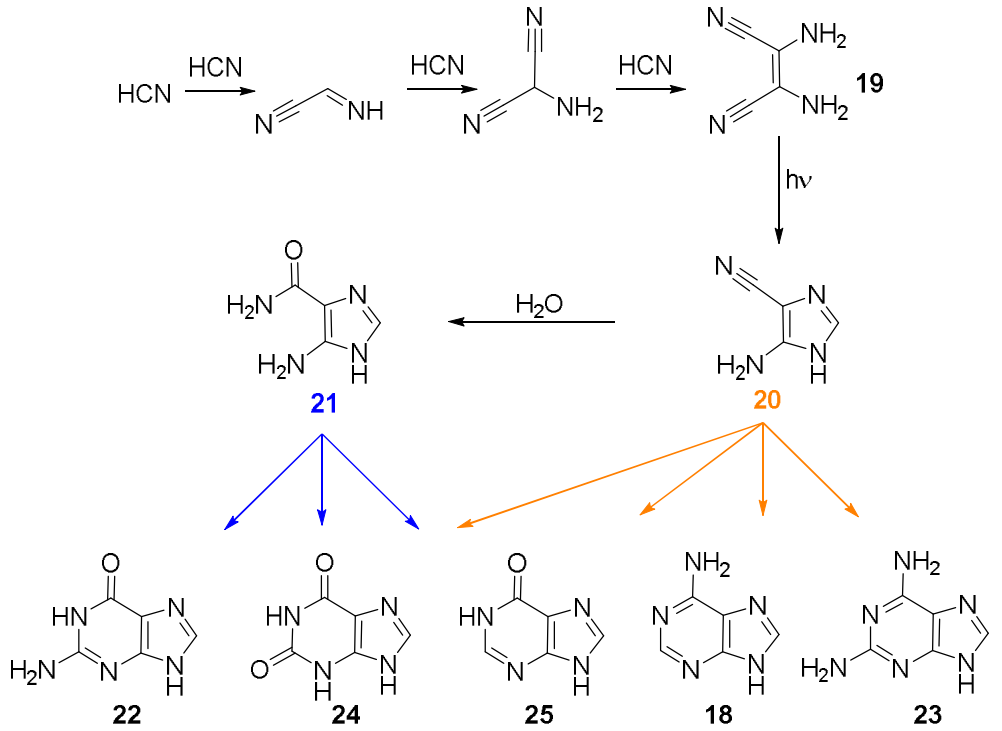
1.2.3. Building nucleobases

The prebiotic availability of the pyrimidine and purine bases has also been studied. The process of prebiotically plausible purine nucleobase synthesis has been described as a "remarkably simple and efficient reaction"[138]. In spite of relatively low yields, what is 'remarkable' in the synthesis of adenine (**18**) is a complex heterocycle can be formed solely from the self-condensation of HCN[191,192]. Constitutionally, HCN is one of the simplest organic compounds, and can easily be formed in spark discharge experiments[44,176,178,193,194]. Oró demonstrated, in 1960, that heating HCN at 70°C, in the presence of ammonia produced 0.5% adenine (**18**)[191,192]. Lowe *et al.* later confirmed the results and

stated that similar conditions, interestingly, were found to produce amino acids as well as other purines[124] (*vide infra*). However, the use of high concentrations of cyanide and ammonia[195,196] (2-3M), as well as the incompatible reaction conditions (initial heating in strong base and subsequent hydrolysis in cool acid) was criticised[166]. Schwartz *et al*. bypassed the need for high concentrations by using the eutectic phases of ice[197], which allow for much lower concentrations of reagents (of the order of 10mM) to be used[ii]. They also showed that far from being incompatible with 'oxygenous' chemistry, this 'nitrogenous' chemistry is actually improved by the addition of aldehydes[197,198] (mechanistic and reaction details are provided in *Section 7*).

Another possible pathway to purines is through the photo-isomerisation of dilute HCN. Ferris and Orgel report efficient formation of up to 15% HCN tetramer diaminomaleonitrile (DAMN; **19**)[199]. Once the tetramer is formed, other purine precursors, such as 4-amino-imidazole-5-carbonitrile (AICN; **20**), can be accessed in good yields through photo-isomerisation (up to 82%)[200] (*Scheme 1.2*). AICN (**20**) hydrolyses to purine precursor 4-amino-imidazole-5-carboxamide (AICA; **21**). Access to purines (adenine (**18**), guanine (**22**), diaminopurine (**23**), xanthine (**24**), hypoxanthine (**25**)) is then obtained through heating of **20** or **21** with a variety of prebiotic compounds such as ammonium cyanide, cyanogen (**5**) and ammonia, or potassium cyanate (**26**). The maximum yield of the canonical purines **18** and **22** are of 11% and 43% respectively (starting from **20** and **21** respectively)[201](*Scheme 1.2*). The simplicity of the synthesis is thwarted by the moderate yields obtained, as well as the insolubility of purine bases in water[202], putting into question their availability in the aqueous medium used in (most) prebiotic chemistry.

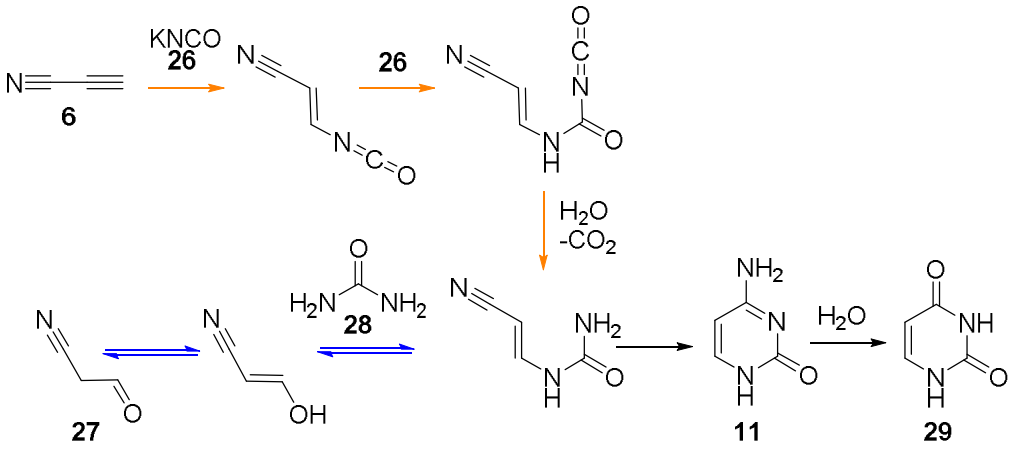
[ii] The low concentrations of HCN would be increased by freezing out the competing solvent, water.



*Scheme 1.2: Proposed prebiotic route to purine nucleobases: synthesis of HCN tetramer diaminomaleonitrile (**19**) from HCN, and subsequent purine synthesis through precursors AICN (**20**) and AICA (**21**), forming adenine (**18**), hypoxanthine (**25**), xanthine (**24**), diaminopurine (**23**) and guanine (**22**).*

Prebiotic pyrimidine syntheses have proved challenging and have often resulted in low yields[203]. Cytosine (**11**) synthesis has been reported by Ferris *et al.* in the reaction of cyanoacetylene (**6**), which is a product of electric discharge on a mixture of methane and nitrogen[203], with cyanate (**26**; *Scheme 1.3*, orange pathway). However, the reaction conditions are quite harsh, involving heating a 0.1M solution of cyanoacetylene (**6**) with 0.3M potassium cyanate (**26**) at 100°C for 24 h to only produce a 7.6% yield. The low yield is no doubt compounded by a fast hydrolysis rate of **26** ($T_{1/2}$ = 43.3 h, 22ºC[204]). A significant improvement was achieved upon reacting cyanoacetaldehyde (**27**; the hydrolysis product of **6**[205]) with concentrated aqueous urea (**28**) at 100°C for 2 h, yielding 30-50% cytosine (**11**)[206] (*Scheme 1.3*, blue pathway). The improbability of such high concentration of **28** and **27** is possibly by-passed by invoking the presence of drying lagoons or pools on the early Earth that could have been home to such an

environment[63]. Once cytosine (**11**) is obtained, uracil (**29**) is easily accessible through simple hydrolysis[203].



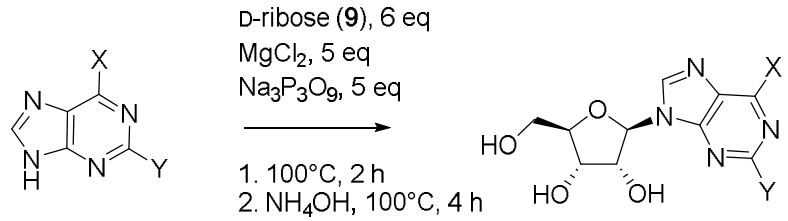
*Scheme 1.3: Proposed prebiotic routes to pyrimidine nucleobases: synthesis of cytosine (**11**) from cyanoacetylene (**6**) and potassium cyanate (**26**) or cyanoacetaldehyde (**27**) and urea (**28**), and subsequent hydrolysis of **11** to uracil (**29**).*

1.2.4. Glycosidation

The prevailing problem in the prebiotic synthesis of ribonucleotides is the selective glycosidation of bases. Indeed, it is tempting to think that simply building ribose (**9**) and bases separately and then condensing them to form ribonucleosides would be the simplest and most obvious way this essential RNA building block would have emerged, but this glycosidation pathway is fraught with both kinetic and thermodynamic obstacles (*vide infra*), causing Sutherland and colleagues to go as far to say that "the traditional retrosynthetic disconnection of RNA equates with chemistry that just does not work!"[207]. This glycosidation step is considered the biggest obstacle that prebiotic chemists need to overcome and the weakest chain in the prebiotic synthesis of RNA[63,165,208].

Fuller, Sanchez and Orgel reported that adenosine (**30**) and inosine (**31**) nucleosides could be formed in low yield, from pure D-ribose (**9**) and the

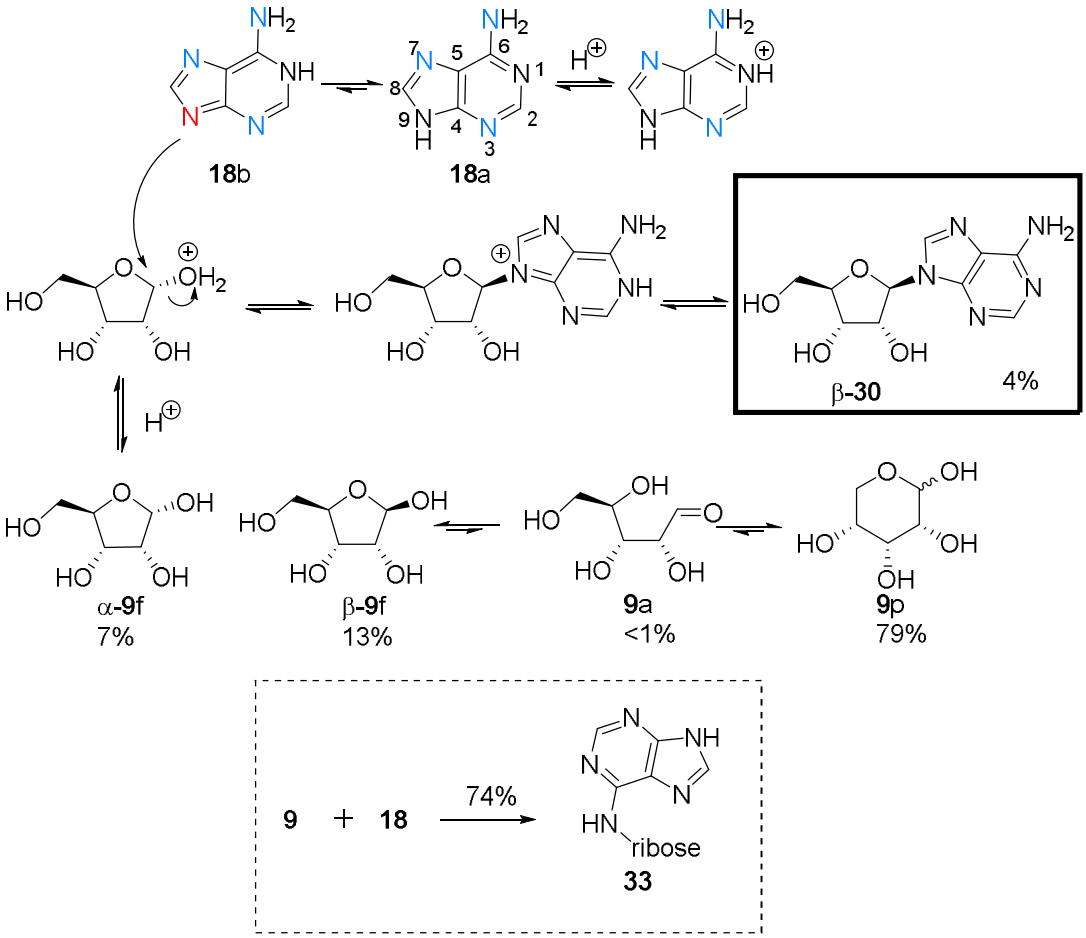
corresponding purine nucleobases. They first screened a variety of conditions: heating (30-100°C), a range of pH (2-11), irradiating with UV light (257 nm and 350 nm), in the presence of prebiotic condensing agents (cyanamide (**4**), urea (**28**), cyanate (**16**), polyphosphates, cyanoacetylene (**6**)) or in the presence of catalysts (minerals, clays or charcoal). No product was detected (detection threshold of 0.1%). They then found that they could obtain small yields of glycosidation to the canonical β-adenosine (β-**30**; ~4%) if they heated adenine (**18**; 0.05M) with an excess of D-ribose (**9**; 6 eq) in the presence of magnesium chloride and sodium trimetaphosphate (5 eq each, both of which are required for the reaction to proceed) at 100°C for 2 h, followed by heating in concentrated ammonium hydroxide solution at 100°C for 4 h[208] (*Scheme 1.4*). Similar conditions were used to obtain β-inosine (β-**31**) in 8% yield. Guanine (**22**) is particularly insoluble in water and thus they adapted the conditions (0.5mM **22** with *300* eq ribose (**9**), and 250 eq of both magnesium chloride and sodium trimetaphosphate) in order to obtain 9% conversion to β-guanosine (**32**)[208].



*Scheme 1.4: Synthesis of purine nucleosides riboadenosine (β-**30**), riboinosine (β-**31**) and riboguanosine (β-**32**) as reported by Fuller et al.[208], when starting with preformed D-ribose (**9**) and the corresponding purine bases (adenine (**18**), hypoxanthine (**25**) and guanine (**22**) respectively). *Reaction conditions differed for the reaction of **22** where 300 eq of **9** were used, along with 250 eq of both magnesium chloride and sodium trimetaphosphate.*

Even if we accept the reaction conditions as prebiotic and accept the synthesis of the starting materials ribose (**9**) and purine bases as predisposed (*vide supra*, *Sections 1.2.2-1.2.3*), the problem remains one of selectivity. Indeed, all other possible anomers and regioisomers are obtained in these reactions, due to the

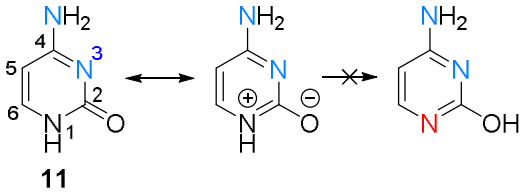
inherent reactivities of ribose (**9**) and adenine (**18**). The sugar **9** is found in multiple forms in water, preferentially in the pyranosyl form (α and β, **9**p), followed by furanosyl (again both anomers, **9**f) and a small amount of open-chain aldehyde is detected (**9**a, through which the latter two can interconvert; *Scheme 1.5*)[209]. Purine nucleobases also exist in various forms in water. For example, in the case of adenine, the major tautomer **18**a, is not reactive through the nitrogen needed for canonical glycosidation, *N*9. This is due to the inaccessibility of the lone-pair of *N*9 needed for nucleophilic attack on the anomeric C1 of ribose (**9**), which is delocalised through the heterocycles[165]. The tautomer of adenine needed for correct reaction, **18**b, is found in very low concentrations in water (*Scheme 1.5*) [165]. The main product obtained is an *N*6-ribosyl adduct **33**, resulting from the reaction of the amino-group of **18** with sugar **9**[208] (*Scheme 1.5*, dashed box).



*Scheme 1.5: Adenine (**18**) glycosidation with ribose (**9**) to form canonical β-riboadenosine (β-**30**; shown in the bold box), and the problems associated with it. Adapted from Sutherland[165], percentages of **9** extracted from Drew et al[209]. Nucleophilic sites of **18** are shown in blue, with the correct nitrogen, N9, shown in red if available for glycosidation. The preferred product of the reaction of **18** and **9**, as performed by Fuller et al., is the non-canonical N6-ribosyl adduct **33** (unreported anomer of ribose isomer; dashed box)[208].*

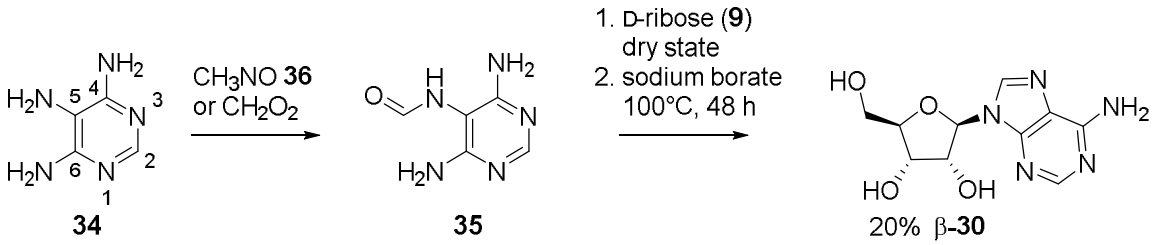
The mixture of products obtained from the reaction of purine nucleobases and ribose (**9**) are then problematic when considering the building of a polymer of RNA, as all these forms could cross-inhibit the oligomerisation and the function of RNA. For example, α-anomeric strands of nucleic acids are considered as a way to regulate mRNA expression *in vivo*[210], due to their propensity to form stronger double strands that are harder to dissociate and may lead to inhibition of the β-anomeric strands.

Attempts to condense the canonical pyrimidine bases with ribose (**9**) have been unsuccessful[165], due to the inaccessibility of the *N*1-nitrogen lone pair involved in the aromatic resonance, making the pyrimidine nucleobases unavailable for glycosidation (*Scheme 1.6*). This is why in conventional synthetic chemistry, glycosidation is achieved by directing the ligation needed by use of protecting and activating groups[211], demonstrating the difficulties to surpass for glycosidation to occur efficiently prebiotically.



*Scheme 1.6: Nitrogen N1 of cytosine (**11**) is involved in aromatic resonance, making it unavailable for glycosidation. Nucleophilic sites of **11** are shown in blue, with the correct nitrogen, N1, shown in red if available for glycosidation.*

As Lazcano and Miller said: "there is no efficient prebiotic synthesis of purine ribosides and no prebiotic synthesis of pyrimidine nucleosides at all."[212] That being said, minor improvements on the selectivity have been obtained through the use of purine precursors triaminopyrimidines (4,5,6-triaminopyrimidine (**34**) or variants), by the Carell group[213]. In addition to improved solubility in water, these purine precursors have the advantage of being symmetrical through the C2-C5 axis, minimising the possible products obtained. Furthermore, they found they could direct the reaction of the correct exocyclic nitrogen (*N*4 or *N*6) to yield a *N*9-riboside with a maximum 20% yield of the canonical furanosyl β-**30** (*Scheme 1.7*; a point to be returned to later, *Section 7*). This strategy does direct selectivity form the perspective of the purine moiety, but doesn't solve the mixture obtained from the ribose moiety (pyranosyl and furanosyl, both α and β), or the source of the pure starting materials needed.



*Scheme 1.7: Improved purine glycosidation to yield canonical riboadenine (β-**30**) via the formylation of 4,5,6 triaminopyrimidine (**34**) to **35** and subsequent dry-state reaction with pure D-ribose (**9**), as reported by Becker et al.*[213]

1.2.5. Phosphorylation

In addition to the problems concerning the synthesis of nucleosides, there are some other fundamental problems concerning the building of RNA, starting with the phosphorylation of ribonucleosides. The phosphate backbone is an essential feature of both RNA and DNA. The question of why phosphate is used so extensively in biology has been asked many times over[214], and some key advantages of phosphate have been brought to light. Indeed, phosphate, when integrated in compounds, can help solubilise them[214], as well as retain them within a lipid bilayer[215]. Additionally, when it comes to its use as a linker in nucleotide polymers, phosphate has the inherent advantage of being trivalent, meaning it can remain ionised even when connected to two nucleotides, or can be hydrolysed to form monomers while remaining stable[214]. The fact that RNA and DNA have that repetitive negative charge in its phosphodiester backbone has been deemed an essential feature that helps maximise its templating potential by limiting its ability to fold (due to charge repulsion)[216]. This templating role is what makes RNA and DNA good informational polymers able to support Darwinian evolution[142,216].

One of the main issues regarding phosphorylation concerns the abundance of inorganic phosphate in aqueous solution on the early Earth. Pasek has shown that phosphorus species such as inorganic phosphate, pyrophosphate, phosphite and hypophosphite could have been formed on the early Earth from the erosion of meteorites containing schreibersite $(FeNi)_3P$[217]. Yamagata *et al.* also proposed

that there would have been some available soluble polyphosphate from volcanic eruptions in the prebiotic broth[218]. The prebiotic availability of phosphate has been put into question due to the low solubility of common phosphate salts, such as the magnesium and calcium salts (solubility of 0.3 mg/100 g and 2 mg/100 g respectively in water at 20ºC). To counter that problem, some research has been done into the use of other phosphorus species such as phosphite[219,220]. Indeed, phosphite, unlike phosphate, is soluble in solutions containing calcium and can similarly be obtained by corrosion of minerals[221]. Phosphite could then be used instead of phosphate in nucleotides in an alternative biochemistry, perhaps preceding RNA[222] (a point to be return to later, *Section 7*).

Nevertheless, if one assumes some degree of inorganic phosphate species are available, the problem remains of how to insert them into nucleosides. Another problem to take into account is one of selectivity of phosphorylation, as canonical RNA and DNA contain a 3′-5′ phosphodiester linkage, which means the monomers need to contain a 5′ or a 3′-(poly)phosphate (a point to be return to later, *Section 1.2.6*)

Phosphorylations of pre-formed nucleosides require phosphate to act as an electrophile, and thus requires acidic conditions ($pK_{a1} \sim 2$). Problematically, when no activating agent is used, the leaving group in the reaction of phosphate and a nucleophile is hydroxide, which with a pK_a of 14, makes it a poor leaving group in acidic conditions. Furthermore, when attempting phosphorylations in water, the hydroxyl groups of the solvent greatly outcompete the hydroxyl groups of the nucleosides[223]. Using prebiotic condensing agents (such as cyanamide (**4**), cyanoformamide (**37**), cyanogen (**5**)), while increasing the electrophilic nature of the phosphate (mechanistic details provided in *Section 3*), does not solve the competition problem, leading to low yields of phosphorylated product (<4%)[223]. One successful aqueous phosphorylation involves the use of cyclic trimetaphosphate in strongly alkaline conditions, yielding 2′ or 3′-phosphorylated nucleotides preferentially and in high yields (90% after 4 days in 1M NaOH)[224]. However, the high pH of the reaction and questionable prebiotic availability of

cyclic trimetaphosphates has put into question the prebiotic relevance of the reaction[63,225].

A lot of work has been carried out on alternative non-aqueous solvents and dry-state reactions. Heating nucleosides with inorganic phosphate in pure formamide (**36**) is successful in yielding good amounts of phosphorylated material (50% phosphorylated products after 15 days, 70°C)[226], however the yield decreases dramatically if the formamide (**36**) is wet (a point to be returned to later, *Section 3*). Efficient phosphorylation is observed if one forgoes solvents all together and the reaction is done in the dry-state[227], with up to 96% phosphorylated material recovered in urea-mediated dry-state phosphorylations using ammonium phosphate[225] (greater detail concerning reaction and mechanism is provided in *Section 6*). Up to 20% yield of phosphorylated material can also be observed in the dry-state phosphorylations using a mineral source of phosphate (hydroxylapatite)[225]. The efficiency of these reactions is however thwarted by the complexity of the mixture: mono- (β-**10**/β-**38**/β-**39**/β-**40**), di- (β-**41**/β-**42**), or cyclic phosphates (β-**43**/β-**44**) in a mixture of C5′, C3′ and/or C2′ phosphorylation patterns[225–228] (*Figure 1.13*). The selectivity for C5′-phosphorylation can be improved, leading to a 2:1 preference for C5′-phosphorylation over C2′ or C3′-phosphorylation by addition of periodate (which binds *cis*-diols selectively, thereby acting as a *de facto* protecting group for the 2′ and 3′)[229].

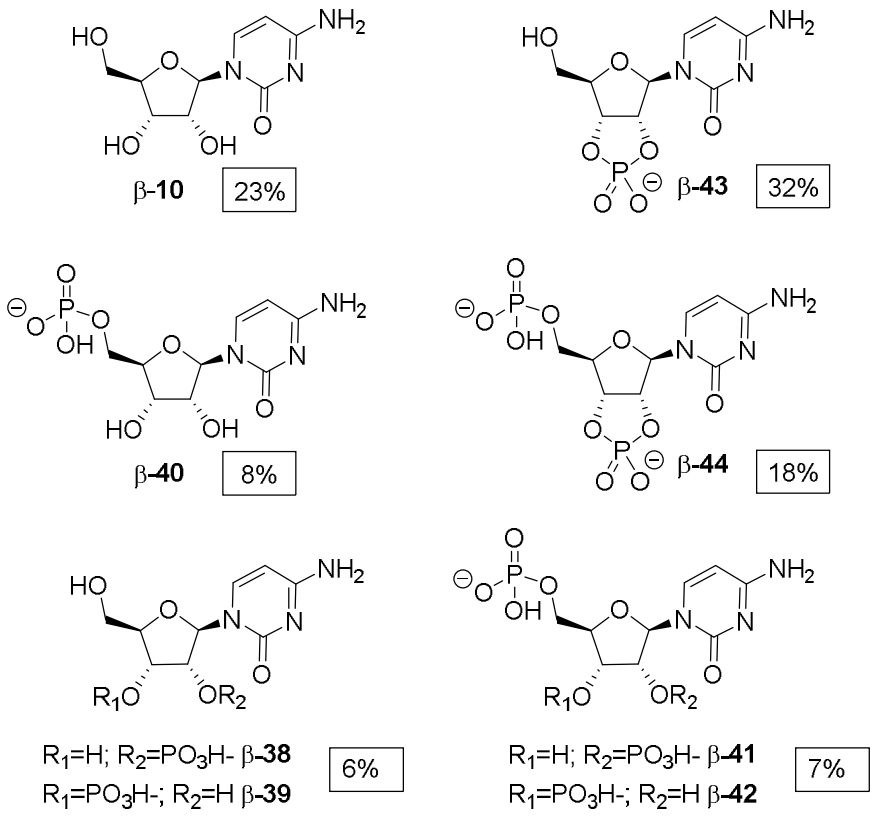
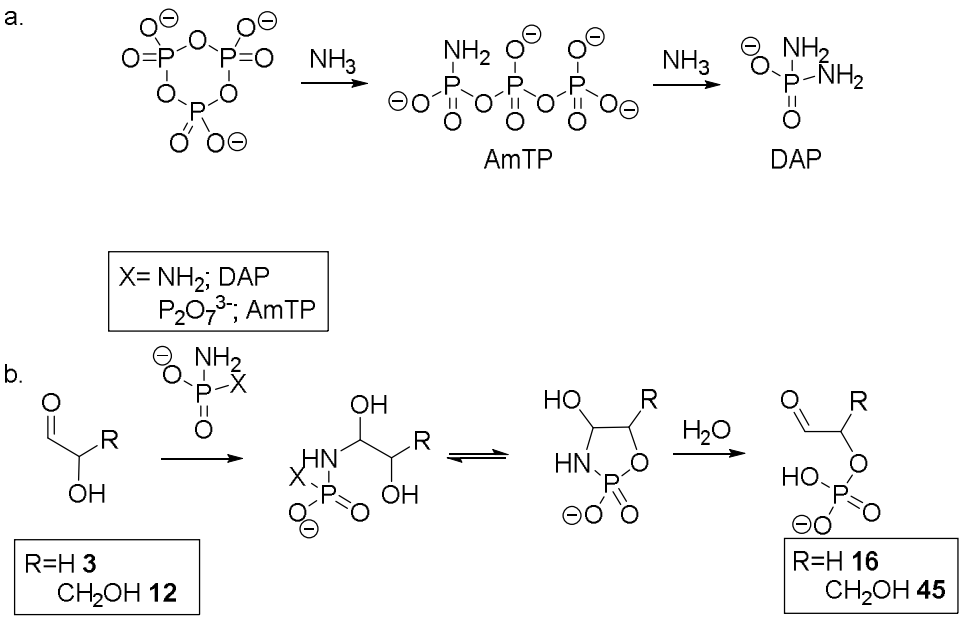


*Figure 1.13: Complex product mixture of the dry-state phosphorylation of β-ribocytidine (β-**10**) with ammonium phosphate in the presence of urea (**28**), after 24 h at 100°C, as reported by Lohrmann and Orgel*[225].

Another way of incorporating phosphate into nucleotides has been through phosphorylation of its precursors. We already mentioned the use of glycolaldehyde-phosphate (**16**) in the synthesis of ribose-2,4-diphosphate (**15**; *Section 1.2.2.*). Glycolaldehyde-phosphate (**16**) and glyceraldehyde-2-phosphate (**45**) can both be formed through reaction with amidotriphosphate (AmTP) or diamidophosphate (DAP; formed from reaction of ammonia and cyclic trimetaphosphate)[189,230] (*Scheme 1.8*). However, as discussed previously, they do not lead to a canonical phosphorylation pattern.



*Scheme 1.8: a. Synthesis of amidotriphosphate (AmTP) and diamidophosphate (DAP) from cyclictriphosphate and ammonia. b. Phosphorylation of α-hydroxyaldehydes glycolaldehyde (**3**) and glyceraldehyde (**12**) with AmTP or DAP.*

The same reaction can be used on D-ribose (**9**) to obtain cyclic phosphates in 71% yield: D-ribose-1,2-cyclicphosphate (**46**) and D-ribose-2,3-cyclicphosphate (**47**; in both the puranosyl and furanosyl forms) with a preference for **46** (43% yield)[189], neither of which have been used for an improved glycosidation reaction. The selective aqueous phosphorylation of D-**9** with phosphate (activated by cyanogen (**5**)) to form β-D-ribofuranosyl 1-phosphate (**48**) in 10-20% yield[231] also suffers from the fact that a non-canonical phosphorylation pattern is obtained.

1.2.6. Oligomerisation and replication

One of the last things to consider in the making of RNA is the assembly of ribonucleotides into oligomers (*Figure 1.14*). If we assume the abundance of building blocks is sufficient and that the prebiotic formation of ribonucleotides is plausible (*vide supra*), we still need to find a credible explanation to the origin of polynucleotides and RNA molecules capable of catalytic activity. One major goal in prebiotic biochemistry is to build, prebiotically, a functioning RNA replicase

(or RNA-dependent RNA polymerase), a ribozyme that could catalyse its own replication[120]. This would bring us closer to "a self-sustaining chemical system capable of Darwinian evolution"[11] (*vide supra*, *Section 1.1.1*.). In order to attempt to build polynucleotides, first activated ribonucleotides must be available for oligomerisation.

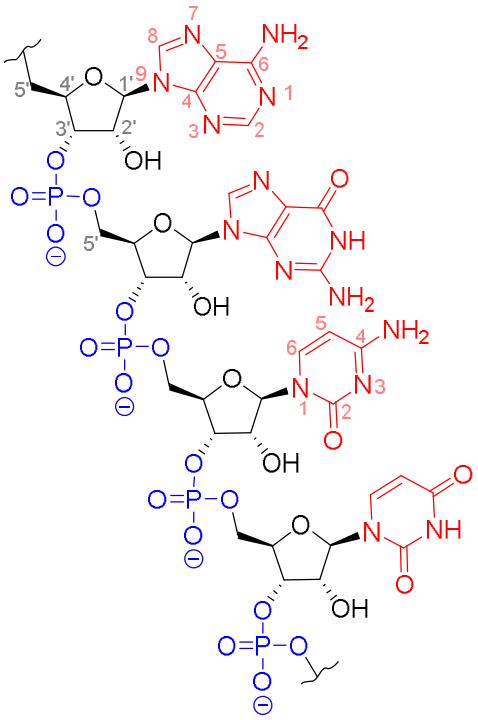


Figure 1.14: Strand of RNA depicted with the four canonical nucleobases (shown in red), highlighting the canonical 3′-5′ phosphodiester linkage (blue) between the sugar backbone (black).

The canonical linkage in RNA/DNA is a phosphodiester bond between the C5′ and C3′ of two nucleotides (*Figure 1.14*). The nucleotide monomers must therefore be 5′ or 3′-phosphorylated, or a 2′,3′-cyclic phosphate. In order to link the monomers, the phosphate first needs to be activated[165]. Nucleotides 2′,3′-cyclic phosphate could arguably be considered "activated" compared to mono-phosphates, but not enough so to feasibly allow a nucleophilic attack by a hydroxyl group[232]. Additionally, oligomerisation reactions attempted with nucleotides 2′,3′-cyclic phosphate lead preferentially to the non-natural 2′-5′ linkages, at a rate comparable to their hydrolysis[233,234]. Attempts to improve the selectivity of the linkages were met with moderate success[235]. Activation of

nucleotides 3′-phosphates leads to cyclisation onto the 2′-hydroxyl group[223,235–237], producing nucleotide 2′,3′-cyclic phosphate, and therefore 3′-phosphates are not usually considered feasible monomers for the oligomerisation of RNA. The plausible/implausible use of 2′,3′-cyclic phosphates in prebiotic context will discussed in greater detail in *Section 6*.

Nucleotide 5′-phosphates have been considered the preferential monomer for oligomerisation due to their use in contemporary biology (nucleotide 5′-triphosphates are used as monomers) as well as the lack of adjacent hydroxyl group onto which an activated phosphate could cyclise. Many conditions have been studied, from the dry-state to the eutectic phase of ice[64,238,239], with clay catalysts[240,241], and various activating agents[240,242–244]. These usually yield a mixture of products with both 3′-5′ and 2′-5′ linkages, with a preference for the natural 3′-5′ linkage[244,245]. Nucleotide 5′-triphosphates, which have the potential to be obtained from nucleotides monophosphates[246], have also been used as activated monomers in a non-enzymatic template-directed ligation where the preference for the canonical 3′-5′ linkage over the unnatural 2′-5′ is up to 80:1[243,247].

These experiments described above have shown that some amount of uncatalysed oligomerisation or ligation can be achieved, although there seems to be a limit on the length that can be produced (<50 nucleotides[248–250]). In order to imagine the rise of a system capable of Darwinian evolution, one needs to show that out of this mixture of short oligomers (formed through templated or untemplated means), sequences could form that inherently have catalytic function (through their secondary structure). Ribozymes capable of improving ligation have been developed through the use of *in vitro* evolution[150,151,251,252], demonstrating proof of concept. The successful design of a ribozyme that can accurately catalyse the synthesis of an RNA sequence longer than itself[253] shows the true potential for the catalytic power of RNA. Finding an RNA sequence that would be capable of replicating its own sequence would cement the concept that evolution from a pool of RNA oligomers is possible, however it has not been considered a requirement. Indeed, one could imagine an autocatalytic set (such as the one described by

Kauffman[254]), where cooperation between the different molecules of the set leads to replication of the whole set. A simplified version involving two ribozymes has shown that their cross-replication could lead to amplification of the desired sequences[255] (*Figure 1.15*). These proof-of-concept experiments have shown that oligomerisation of monomers and ligation of short oligomers can lead to the development of catalytic activity that has the potential for Darwinian evolution, and as Higgs and Lehman said "the evolution of replication probably proceeded in stages."[95]

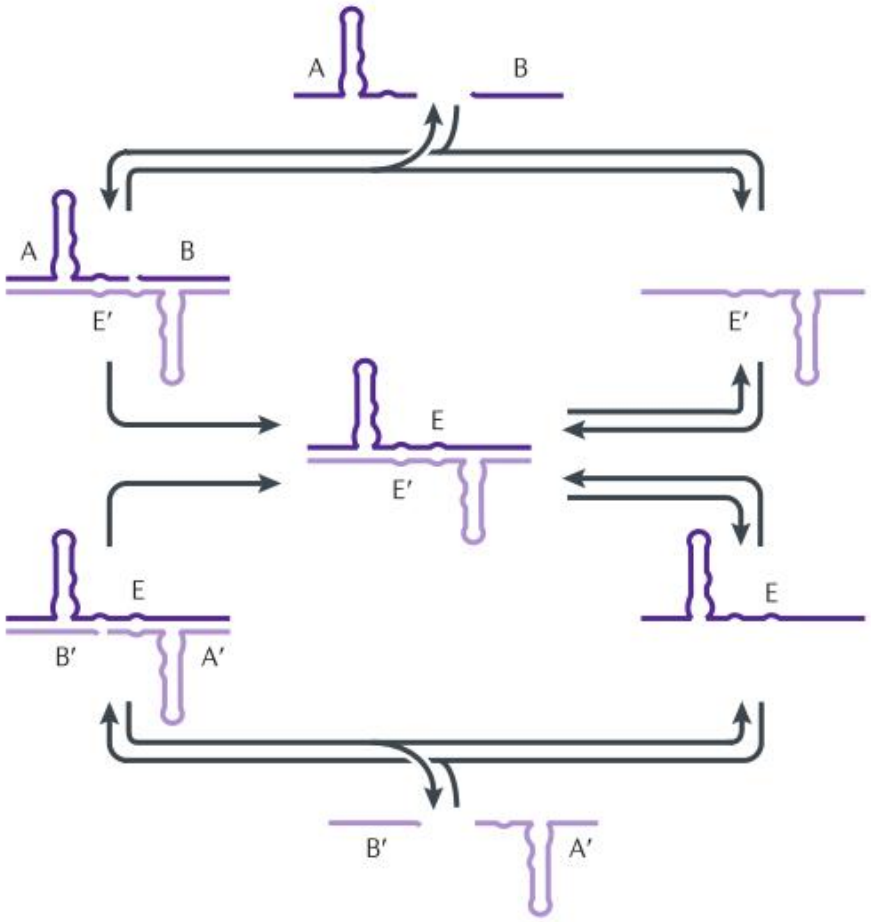


Figure 1.15: Autocatalytic set through cross-replication of ribozymes E and E'. Ribozyme E' catalyses the ligation of RNA A and B to form ribozyme E. Ribozyme E catalyses the ligation of RNA A' and B' to form ribozyme E'. From Higgs and Lehman[95], originally adapted from Lincoln and Joyce[255].

Although many problems still need to be overcome, particularly concerning the synthesis of RNA monomers, the centrality of RNA in the evolution of life has some strong experimental backing.

1.3. Alternative approaches

1.3.1. Pre-RNA world theory

The problems associated with prebiotic RNA synthesis, has led a number of groups to postulate that non-endogenous nucleic acid systems may have preceded RNA and DNA. During an extensive study of the constitutional neighbourhood of RNA, the Eschenmoser lab found that a constitutional isomer of RNA, the pyranosyl form of RNA (p-RNA; *Figure 1.16*) [234,256,257], had a number of advantages relative to conventional RNA.

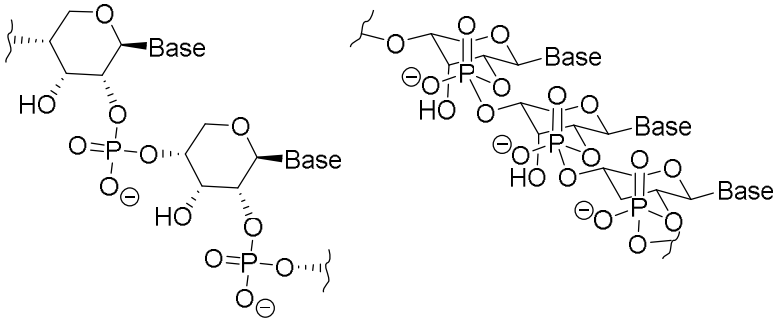


Figure 1.16: Structure of pyranosyl RNA (p-RNA), where base represents a nucleobase.

Firstly, they had previously demonstrated that they could form the phosphorylated sugar moiety, ribose-2,4-diphosphate (**17**) selectively when starting with glycolaldehyde-phosphate (**16**) and formaldehyde (**2**)[187] (*Section 1.2.2*). Secondly, p-RNA, while still being able to form Watson-Crick base pairs and duplex like RNA, is less likely to form multiple-strand competing structures. Lastly, the base pairing is stronger, more stable and more selective than in RNA. Although higher duplex stability is beneficial for genetic integrity, it disfavours the strand separation needed for replication. The downside to the p-RNA theory is that it

cannot pair with RNA, which would make direct "genetic take-over" by the endogenous system highly improbable[257].

Other "pre-RNA" biopolymers have been proposed, such as peptide nucleic acid (PNA). PNA is an achiral, uncharged analogue of RNA where the nucleobases are linked by peptide bonds[258] (*Figure 1.17*). PNA has received a lot of interest due to its ability to form very stable double helices with both DNA and RNA with the correct Watson-Crick base pairing[259]. Studies by Schmidt *et al.* have shown that sequence information can be shared between DNA and PNA[260], allowing, in theory, for an easy "genetic take-over" from PNA by the endogenous polynucleotides, something that is lacking in the p-RNA theory mentioned above[257]. However, it has yet to be demonstrated that PNA monomers can be synthesised or that acyclic monomers could polymerise (instead of an expected rapid cyclisation)[63].

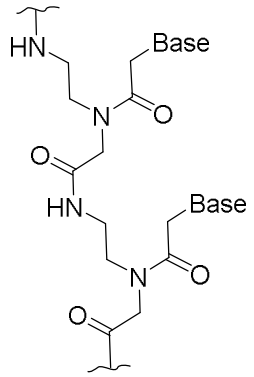


Figure 1.17: Structure of peptide nucleic acid (PNA), where base represents a nucleobase.

Threose nucleic acid (TNA) is of particular interest as it is constituted of a simpler sugar backbone (only four carbon atoms) and the prebiotic synthesis of the tetrose sugars have been demonstrated by the Eschenmoser group[261] (*Figure 1.18*). Furthermore, like PNA, TNA is able to form complementary strands with RNA and DNA, and in spite of having a "shorter" backbone (5 atoms instead of the 6 found in RNA/DNA) it can adopt a stretched conformation where the substituents on the furanose ring are quasi-axial[262]. Theoretically, this could allow for a

"genetic take-over". However, this constitutional simplicity of TNA means it has a reduced phenotypic capacity by virtue of the loss of the equivalent of RNA's C2′ hydroxyl group[262].

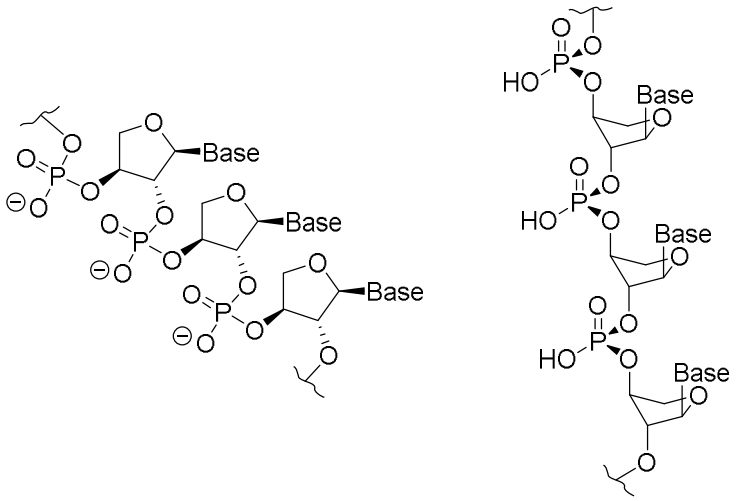


Figure 1.18: Structure of threose nucleic acid (TNA), where base represents a nucleobase.

Other proposed variants include glycol nucleic acid (GNA), that is also constitutionally simpler than RNA, can base-pair with RNA, but with no known prebiotic synthesis of its monomers[263] (*Figure 1.19*, *a.*); and alanyl nucleic acid (ANA), a polypeptide formed from a racemic nucleo-amino acids mixture[264,265] with alternating chirality (*Figure 1.19*, *b.*).

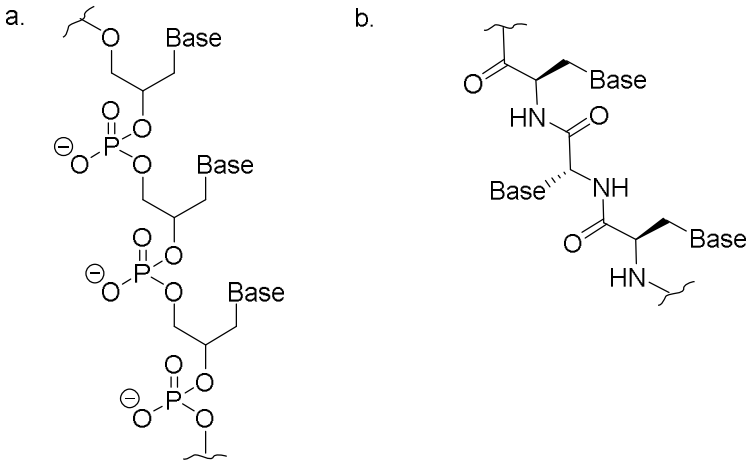


Figure 1.19: Structures of a. glycerol nucleic acid (GNA) and b. alanyl nucleic acid (ANA), where base represents a nucleobase.

Despite the lack of any "hints" from biology, that could help direct the search of potential pre-RNA polymers, these xeno nucleic acids (XNA) have been fruitful in providing different options, with more or less prebiotic weight. However, invoking a pre-RNA polymer doesn't really simplify the question of the evolution of RNA, and instead raises new ones: How did *that* polymer evolve? Why did biology not keep it as the main biopolymer? Why is there no trace of it in biology? Or as Sutherland asked: "Nature has a habit of changing the function of newly redundant entities rather than disposing of them – wouldn't there then be some traces of XNA in extant biology?"[38].

As Cairns-Smith stated, the problem isn't only in the prebiotic synthesis of the building blocks of life but in "the spontaneous appearance of information"[266]. He proposed a "genographic" origin of life, where the emergence of hereditary information essential to life could have been found in inorganic systems, such as clay crystals[266]. He postulates that crystal "replication" (growth) and "evolution" (imperfections that change the crystal lattice or its chemical composition) could initiate a breach through the information barrier, leading to a primitive genetic information system. Another interesting theory is depicted in the "triose model", where glyceraldehyde (**12**) provides a (theoretical) link between metabolism and polymer synthesis (in the form of the proposed polyglyceric acid)[267]. This model remains without experimental validation, but does benefit from attempting to link areas of biology and origin of life studies that are traditionally disconnected. As we will discuss in the following sections, such links between areas of biology (metabolism, biopolymers, membrane, etc) do exist, and research into the origin of life can benefit from connecting previously distinct areas of prebiotic research.

1.3.2. Systems chemistry and recent advances

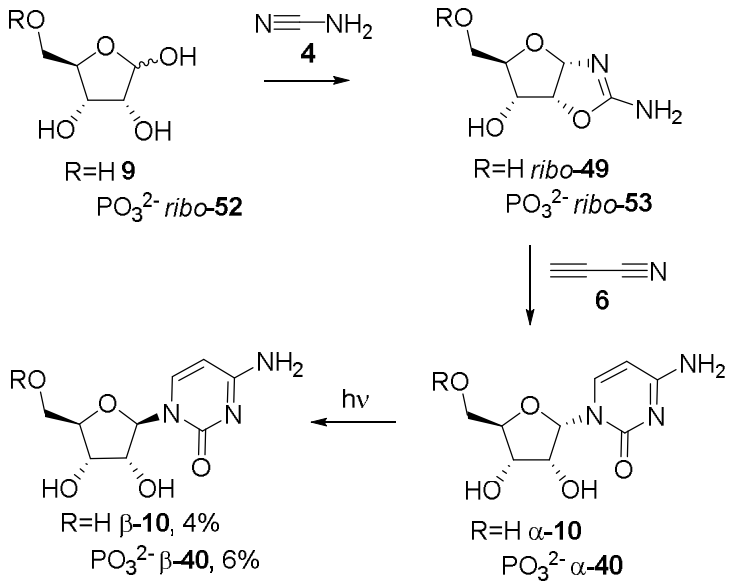
Through consideration of the known prebiotic chemistry of RNA, Shapiro came to the conclusion that "the spontaneous formation of even a short nucleic acid chain under prebiotic conditions was a highly improbable event. It is likely that nucleic acids were first assembled with the aid of enzymes, and are later products of

biochemical evolution"[166]. But perhaps the question of how life originated has been approached in the wrong way. It is clear from what we know of early geochemistry (*Section 1.1.2*) that the Earth would have been abundant in a variety of simple and more complex molecules. Yet the route to the origin of life was always approached as such: a route, a straight line. Separating the different types of chemistry, nitrogenous and oxygenous, under the assumption that they would negatively interfere with each other[190], leading to a "combinatorial explosion" of unwanted by-products, led to a seaming dead-end in the search for the origin of RNA. Separating the different constituents of biology (membrane, metabolism, proteins, nucleic acids, etc.) has led to the view that one is more important than the other and that they can't be studied concomitantly. Advances in analytical tools, and the blossoming of nearby fields like systems biology, has led to a fresh look on some old chemistry: "systems chemistry"[135,136,268,269].

> *"By focusing not on single reactions in isolation but on the collective set of processes that must all occur contemporaneously, systems chemistry can lead to the discovery of 'multifers' in which the system as a whole can establish itself through the shared use of common conditions, reactants or products"* - Higgs and Lehman[95]

Tremendous advances have been made by this opening of the reaction space. Some hints that nitrogenous and oxygenous chemistries might not be as incompatible as they were thought had been seen even in the "traditional" synthesis of nucleobases, as aldehydes were found to catalyse the oligomerisation of HCN to form purine bases (*vide supra*, *Section 1.2.3*)[197,198]. Another early "timid" mixed chemistry was developed by Sanchez and Orgel, in order to find a way around the, as of yet impossible, glycosidations of pyrimidine nucleobases[165,212] (*vide supra*, *Section 1.2.4*). Indeed, they found that by reacting cyanamide (**4**) with ribose (**9**), in the presence of ammonia, an exothermic reaction ensued yielding ribofuranosyl aminooxazoline (*ribo*-**49**) in very good yield (87% yield upon crystallisation with methanol at 4°C)[270]. Subsequent reaction of *ribo*-**49** with cyanoacetylene (**6**; 60°C, pH 8.5) yielded ribofuranosyl-

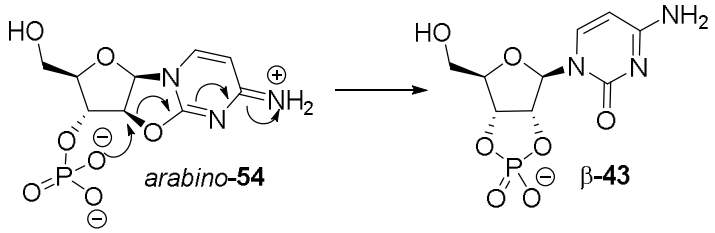
2,2′-anhydrocytidine (*ribo*-**50**) which then hydrolyses under the reaction conditions to yield 64% α-ribofurnasyl cytidine (α-**10**; *Scheme 1.9*). A similar synthesis can be achieved when starting with arabinose (**13**), resulting in the formation of β-arabinofuranosyl cytidine (β-**51**). Although this prebiotically plausible synthesis avoided the problematic glycosidation step, it led to the formation of cytidines that were one stereochemical inversion away from the canonical β-ribofuranosyl cytidine (β-**10**), in the C1′ and C2′ position respectively. Sanchez and Orgel now needed to find a way to anomerise the C1′ of α-**10** to β-**10** or epimerise the C2′ of β-**51** to β-**10**. The first was attempted through photochemistry. They found that by irradiating a solution of α-**10** for 16 h (or the C5′ phosphorylated variant ribocytidine-5′-phosphate (α-**40**) that can be obtained from ribose 5-phosphate (*ribo*-**52**) through the same two-step synthesis as described above), they could obtain low yields of anomerisation of the order of 4% (or 6% in the case of α-**40**; *Scheme 1.9*). Other later photoanomerisation efforts will be discussed in *Section 6*.



*Scheme 1.9: Bypassing the problematic glycosidation by building the pyrimidine nucleobase step-wise on ribose (**9**) or ribose 5-phosphate (ribo-**52**).*

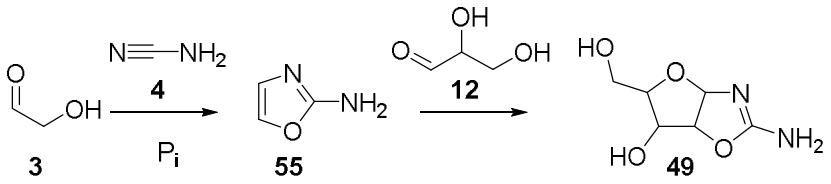
In the case of C2′ inversion in the *arabino*- variant, a solution was found in the form of a phosphate-mediated intra-molecular rearrangement of

arabinofuranosyl-2,2′-anhydrocytidine 3′-phosphate (*arabino*-**54**) to β-ribocytidine-2',3'-cyclic phosphate (β-**43**), discovered by Tapiero and Nagyvary[271] (*Scheme 1.10*). The problem in all of these syntheses, however, is the source of sugar or sugar-phosphate. The synthesis of ribose (**9**), additionally to being low yielding (*vide supra*, *Section 1.2.2*), is also disrupted by the presence of nitrogenous compounds such as cyanamide (**4**)[166].



*Scheme 1.10: Intramolecular nucleophilic displacement in arabino furanosyl-2,2'-anhydrocytidine 3'-phosphate (arabino-**54**), to form β-ribocytidine-2',3'-cyclic phosphate (β-**43**).*

In an attempt to solve this problem, the Sutherland group postulated, through a retrosynthetic analysis, that the problem could be solved through the use of truly mixed chemistry. They proposed that instead of building the pentose sugar moiety first and then building the pyrimidine base in stages onto that pre-formed sugar, one could form a sugar/base synthon, 2-aminooxazole (**55**), through the reaction of glycolaldehyde (**3**) and cyanamide (**4**). As the sugar moiety of ribose is composed of 5-carbons (accessible through C_3+C_2 or C_2+C_2+C_1 units), **55** could provide the C_2 unit, with glyceraldehyde (**12**) providing the remaining C_3 unit (reminiscent of the ribose (**9**) synthesis explored by Ricardo *et al.*[184], *Section 1.2.2*). The reaction of **3** and **4** in phosphate buffer led the efficient formation of **55** in high yield (>80% yield)[33]. Adding **12** at this stage led to the synthesis of *rac*-pentose aminooxazolines (**49**) in good yield (~50% yield over the two steps; *Scheme 1.11*), with a preference for the *ribo*- and *arabino*- diastereomers (78% of **49**, combined), both of which sit exclusively as the furanosyl form (*lyxo*-**49** is the only aminooxazoline that has a preference for the pyranosyl form, in a 5:1 ratio[272]).

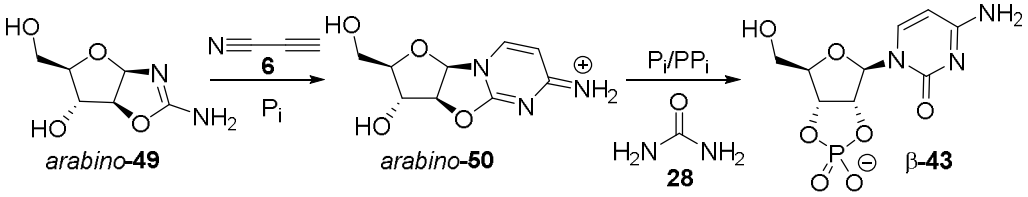


*Scheme 1.11: Synthesis of aminooxazolines **49** through phosphate mediated reaction of C_2 sugar glycolaldehyde (**3**) with cyanamide (**4**) to form 2-aminooxazole (**55**) and subsequent reaction with C_3 sugar glyceraldehyde (**12**); bypassing the need for preformed ribose (**9**).*

This efficient and selective access to aminooxazolines **49** showed the benefits of combining nitrogenous and oxygenous chemistry, by-passing the need for a problematic ribose (**9**) synthesis, and amplifying the *ribo*- selectivity (*ribo*-configuration is 44% of the pentose **49** compared to the 30% ribose (**9**) obtained in the cross-aldol reaction of **3** and **12**[186] or 22% in the aldomerisation of glycolaldehyde phosphate (**16**)[273]). The nature of aminooxazolines **49**, with their obligate 1′,2′ *cis* stereochemistry means that once again the products obtained were one stereochemical inversion away from the canonical nucleotides/nucleosides. Furthermore, the problem of the incorporation of phosphate remained.

Inspired by the Tapiero and Nagyvary rearrangement[271], the Sutherland group set out to find conditions were they could obtain arabinofuranosyl-2,2′-anhydrocytidine 3′-phosphate (*arabino*-**54**) prebiotically, starting with their new found route to aminooxazolines **49**. Firstly, they found that their **49** mixture could be enriched in the *arabino*- variant thanks to the selective crystallisation of *ribo*-**49** upon cooling[272]. Next, they found that by repeating the reaction described by Sanchez and Orgel[270] (*vide supra*) of *arabino*-**49** with cyanoacetylene (**6**), this time in the presence of phosphate buffer, they could now obtain arabinofuranosyl-2,2′-anhydrocytidine (**50**; *Scheme 1.12*). **50** was previously thought of as an intermediate species *en route* to cytidine (**10**). However, it transpired that the hydrolysis of **50** to **10** *in situ* was due to a rise in the reaction's pH, which could be avoided with the use of phosphate as a pH buffer (at pH 6.5). The last step needed to link this synthesis to the C2′ inversion

observed by Tapiero and Nagyvary[271], was to find a way to selectively phosphorylate the C3′ position of *arabino*-**50**. Again, literature precedents of prebiotic phosphorylations (*vide supra*, *Section 1.2.5*) provided some feasible options, such as the efficient, but unselective, urea-mediated dry-state phosphorylation of ribouridine (**56**) reported by Lohrmann and Orgel[225] (*vide supra*). Repeating the reaction with *arabino*-**50** showed a preference for the canonical β-ribocytidine-2',3'-cyclic phosphate (β-**43**; 32% yield or 46% yield when taking into account bis-phosphorylated β-ribocytidine-2′,3′cyclic-5′-bisphosphate (**44**); *Scheme 1.12*). This suggested that the urea-activated phosphate initially phosphorylated the C3′-OH or C5′-OH. This reversible phosphorylation is followed by an (irreversible) intramolecular nucleophilic substitution in the case of C3′-phosphorylation, similar to the one observed by Tapiero and Nagyvary (similar results were obtained with pyrophosphate)[33]. These experiments will be returned to in *Section 6*, along with greater experimental rationale and mechanistic detail.

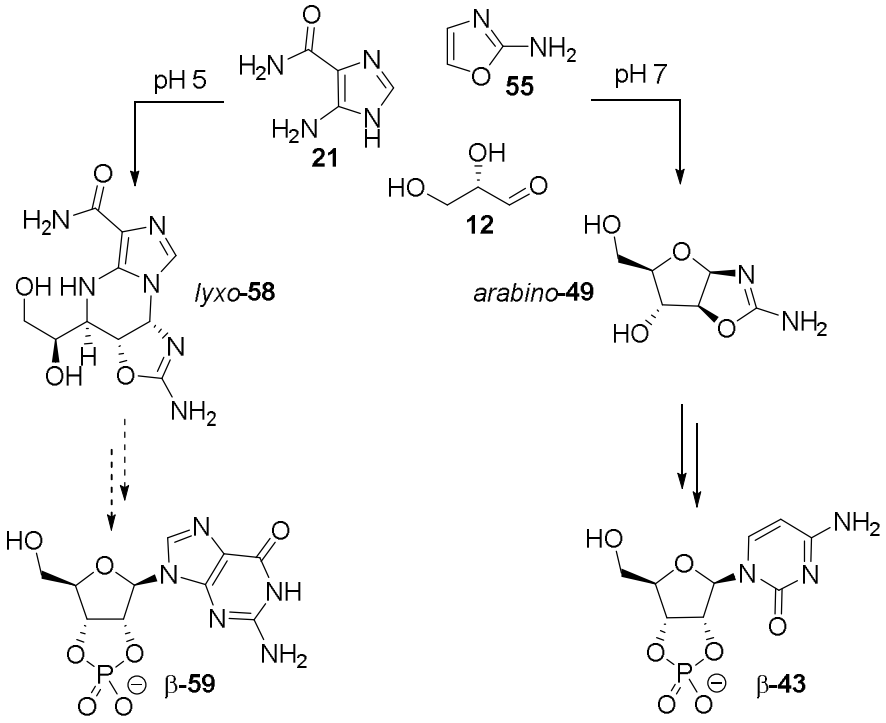


*Scheme 1.12: Synthesis of β-ribocytidine-2',3'-cyclic phosphate (β-**43**) via arabinoaminooxazoline (arabino-**49**), arabino-2,2'-anhydrocytidine (arabino-**50**) and subsequent dry-state, urea-mediated, phosphorylation leading to C2' stereo-inversion.*

The selective synthesis of β-**43** was completed by an irradiation step, that provided a prebiotic way of "cleaning-up" the reaction by selectively retaining the stable β-**43** while destroying other unwanted side products. The irradiation also provided a way of accessing the second pyrimidine β-ribouridine-2',3'-cyclic phosphate β-**57** through partial hydrolysis of the base moiety[33].

The value in this new way of prebiotically synthesising ribonucleotides is that it completely bypasses the need to build the sugar and base moieties separately by using 2-aminooxazole (**55**) as both a sugar and a nucleobase synthon. Furthermore, this synthesis benefits from phosphate as a pH and chemical buffer and general acid/base catalyst in many steps increasing the advantage of the overall system, in line with a systems chemistry approach (*vide supra*). There are still limits to this new synthesis, such as the need for high concentrations of specific materials without competition from closely related compounds[167], the use of dry-state heating in order to incorporate phosphate and step-wise addition of materials. Even with those problems in mind, it is still a great achievement with regards to prebiotic synthesis of pyrimidine ribonucleotides, thus lifting the prejudice against *de novo* synthesis of RNA in prebiotic conditions. However, to realise a pool of monomers capable of generating information-rich RNA, or the "molecular biologist dream"[164] there is still need for a prebiotically plausible (and ideally concomitant) activated purine ribonucleotide synthesis.

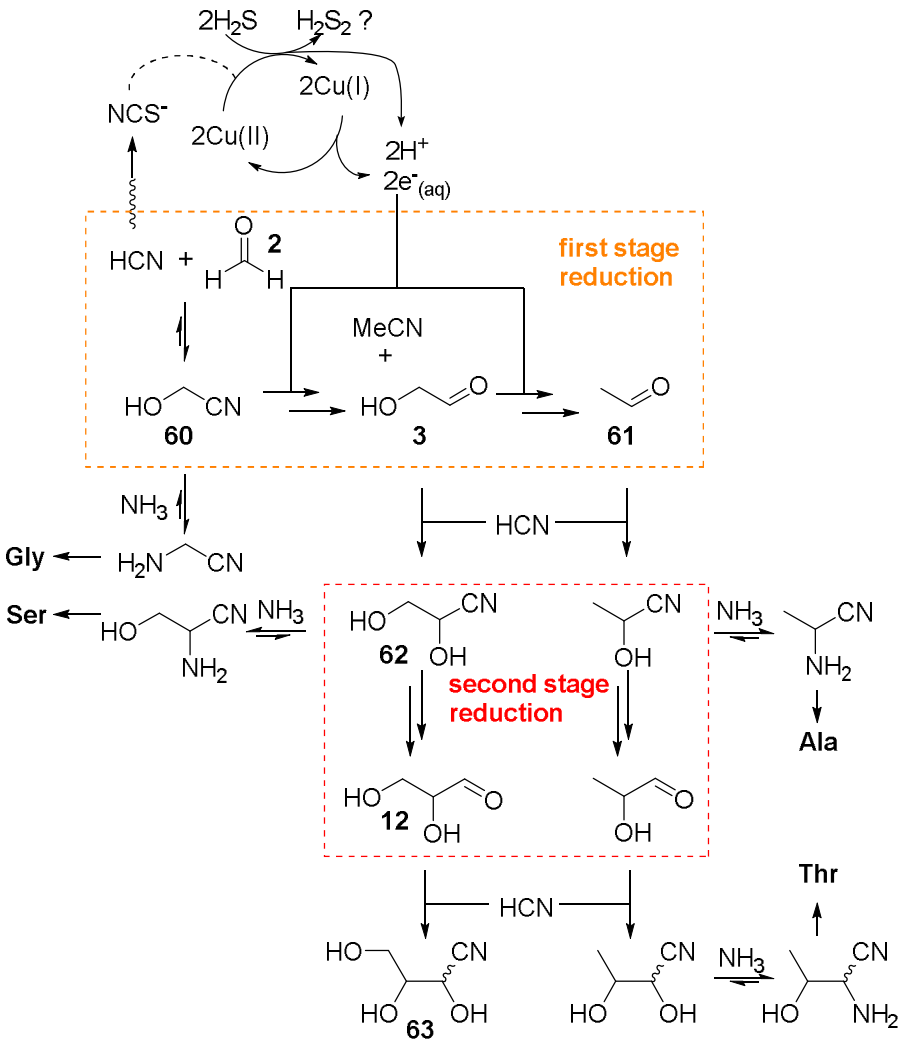
Initial results have been obtained by Powner *et al.*[274], suggesting a concomitant, yet chemically divergent one-pot synthesis of aminooxazolines **49** (leading to pyrimidines) and potential purine precursors. The synthesis developed uses the same prebiotic feedstock materials as described above (**55** and **12**) with the addition of purine base precursors AICA (**21**; which can be obtained through the oligomerisation of HCN, *Section 1.2.3*), with the pH of the solution providing the chemical control of the product distribution (high pH leads to the biomolecular reaction of **55** and **12** to **49**, while low pH leads to the selective synthesis of the 3-component product **58**; *Scheme 1.13*; a point to be returned to later, *Section 7*).



*Scheme 1.13: One pot chemically divergent synthesis of pyrimidine precursors pentose aminooxazolines (**49**) and potential purine precursors **58**. The intermediates shown depict those leading to the natural stereochemistry. Dashed lines represent hypothetical synthesis, while solid lines represent known synthesis.*

The most recent work by the Sutherland group really epitomises the potential of systems chemistry[37,38,275,276]. They demonstrated that sugar feedstock molecules used in the above syntheses, glycolaldehyde (**3**) and glyceraldehyde (**12**), could be obtained from HCN through a photoredox system using cyanocuprates[275], or through photoredox chemistry of glycolonitrile (**60**; obtained from formaldehyde (**2**) and HCN) using hydrogen sulfide as the reductant[276]. HCN is particularly interesting from a prebiotic chemistry point of view, due to its constitutional simplicity and prebiotic availability. HCN is supposedly prebiotically abundant as it can form in spark discharge experiments[193], is widespread in the solar system and comets[277,278] and could have been delivered through meteorites[279]. Furthermore, it is used in a wide variety of prebiotic reactions such as the synthesis of adenine (**18**)[200] and amino acid synthesis through Strecker synthesis[36,280]. In the photoredox system using hydrogen sulfide, Ritson and

Sutherland obtain the sugars **3** and **12** as well as precursors for the extant amino acids glycine (**Gly**), serine (**Ser**), alanine (**Ala**) and threonine (**Thr**; *Scheme 1.14*)[276].



*Scheme 1.14: Photoredox systems chemistry of glycolonitrile (**60**) with hydrogen sulfide, adapted from Riston and Sutherland*[276].

This photochemistry compelled the authors to say that this system shows "hints of a linked origin of ribonucleotides and amino-acids"[276], and shows the necessity to look at the 'bigger picture' when trying to understand the origins of life. Expansion of this photoredox chemistry led to parts of the 'bigger picture' being

assembled, thanks to the prebiotic synthesis of amino acid precursors, nucleotide precursors, membrane precursors, all stemming from this one "cyanosulfidic protometabolism"[37] (a point to be returned to, *Section 2*).

1.3.3. Deciphering the natural system

Through the results obtained in recent years in the field of systems chemistry (*vide supra*), we can see that there is tremendous benefit in forgoing the traditional separation of different areas of prebiotic research into the origin of life in favour of a more comprehensive and integrated view. In this way we can combine some geochemical characteristics of the early Earth while designing experimental goals, in order to see how one affects the other.

> *"Surely, it makes more sense to first find predisposed chemical routes to molecules of interest and then ask whether the sequence of conditions is geochemically plausible."*- Sutherland[165]

We will therefore seek, in this research programme, to broaden the stock of prebiotically plausible molecules that can give rise to a range of prebiotic compounds such as amino acids, pyrimidine nucleotides and purine nucleotides, in a compatible environment. Furthermore, we will attempt to provide a solution to the aforementioned problems still present in pyrimidine nucleotide synthesis, namely the synthesis of 5′-phosphorylated nucleotides in a mild and aqueous environment, by looking at multi-component chemistry, in line with a systems approach to chemistry.

> *"Every experimental research project with the goal of demonstrating a "prebiotic" way of forming RNA or a potential precursor is a contribution to the development of a file of experimental facts, familiarity with which will be prerequisite toward any decisive step in the search for the chemistry of the origin of life"* - Eschenmoser[32]

Finally, any prebiotic research has the potential to uncover other important prebiotic reactions and we will strive to put every result in context of the ultimate goal: trying to understand the chemistry leading to the origin of life.

> *"The scientific search for the route nature adopted in creating the life we know will arguably never truly end. It is, after all, part of the search for our own origin"* - Eschenmoser[32]

2. The use of acrolein in prebiotic synthesis

2.1. Prebiotic synthesis of amino acid precursors

Finding a way to synthesise amino acids and nucleic acids without the use of biological machinery has been central to the study of the origins of life[36,165] (*vide supra*, *Section 1.1.3-1.2*). Seminal work by Miller, under the supervision of Urey, demonstrated production of amino acids from spark discharge experiments on an anaerobic mixture of gases (including NH_3, CH_4, H_2O and H_2), mimicking the influx of energy (such as lightning) on a model of the contemporary early Earth atmosphere[36,281]. The amino acids in the spark discharge experiments are supposedly accessed through Strecker synthesis[280] (*Figure 2.1*) on aldehydes produced *in situ*. Strecker synthesis involves the formation of an imine through the reaction of an aldehyde with ammonia (the latter of which is present in the atmospheric model used in the experiment), followed by nucleophilic attack of HCN, formed *in situ*[176,178,193,194], to form α-amino nitriles (*Figure 2.1*, orange box). The equilibrium lies in favour of aminonitriles over cyanohydrins when the reaction is above pH 8-8.5[282]. Amino acids are then accessed *in situ* from amino nitriles through aqueous acid hydrolysis[iii] (*Figure 2.1*, blue box).

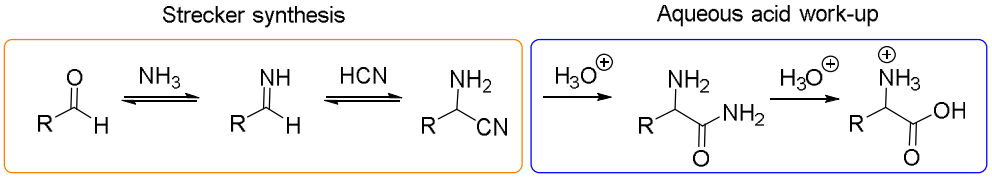


Figure 2.1: Prebiotic synthesis of amino acids via Strecker synthesis on aldehydes, as proposed for the synthesis of amino acids from Miller-Urey type spark discharge experiments.

Amino acid synthesis *via* Strecker synthesis is an essential pillar of prebiotic chemistry that is dependent on what aldehydes are prebiotically available. Through retrosynthetic analysis of the Strecker reaction (*Figure 2.2*), one can

[iii] Hydrolysis can also be achieved in alkaline conditions.

determine the aldehyde precursors needed to access the 20 canonical amino acids (used in biology; *Section 1.1*).

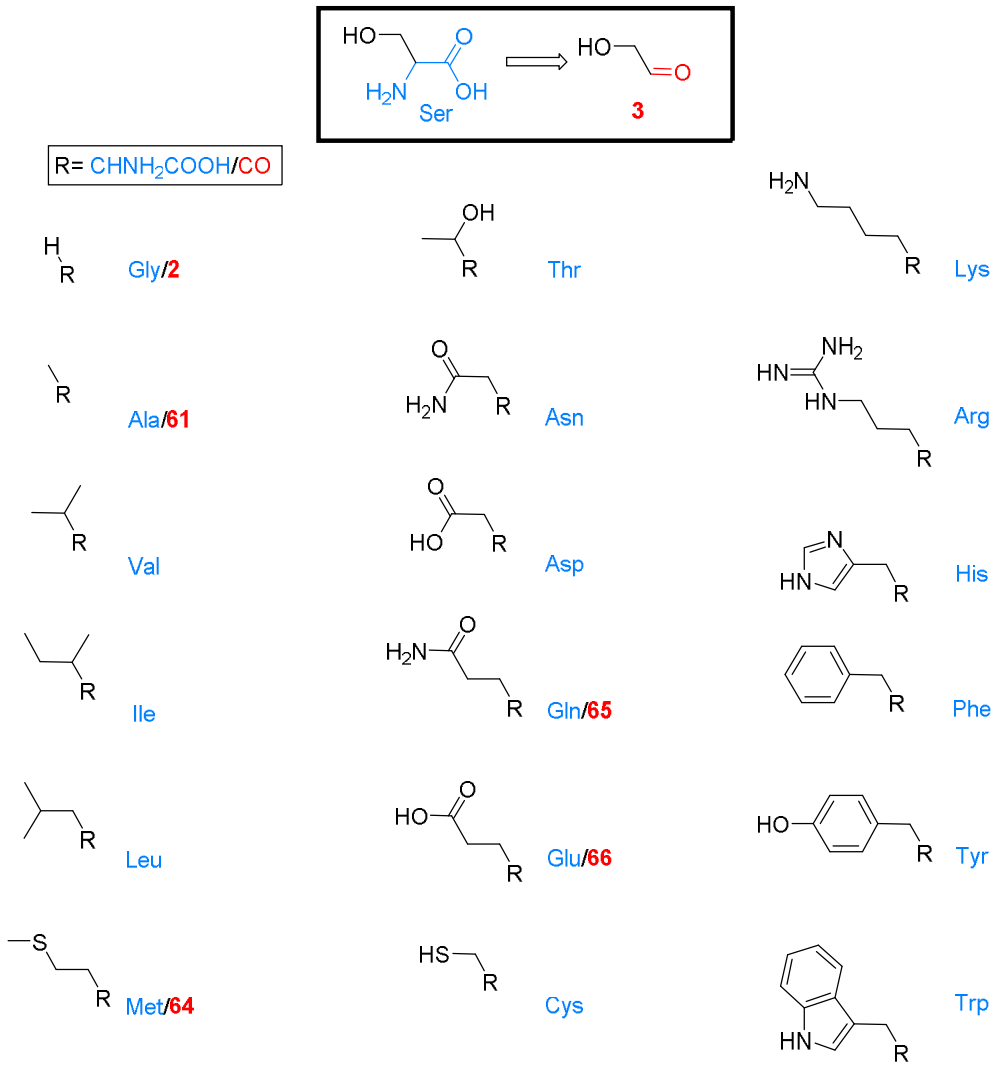
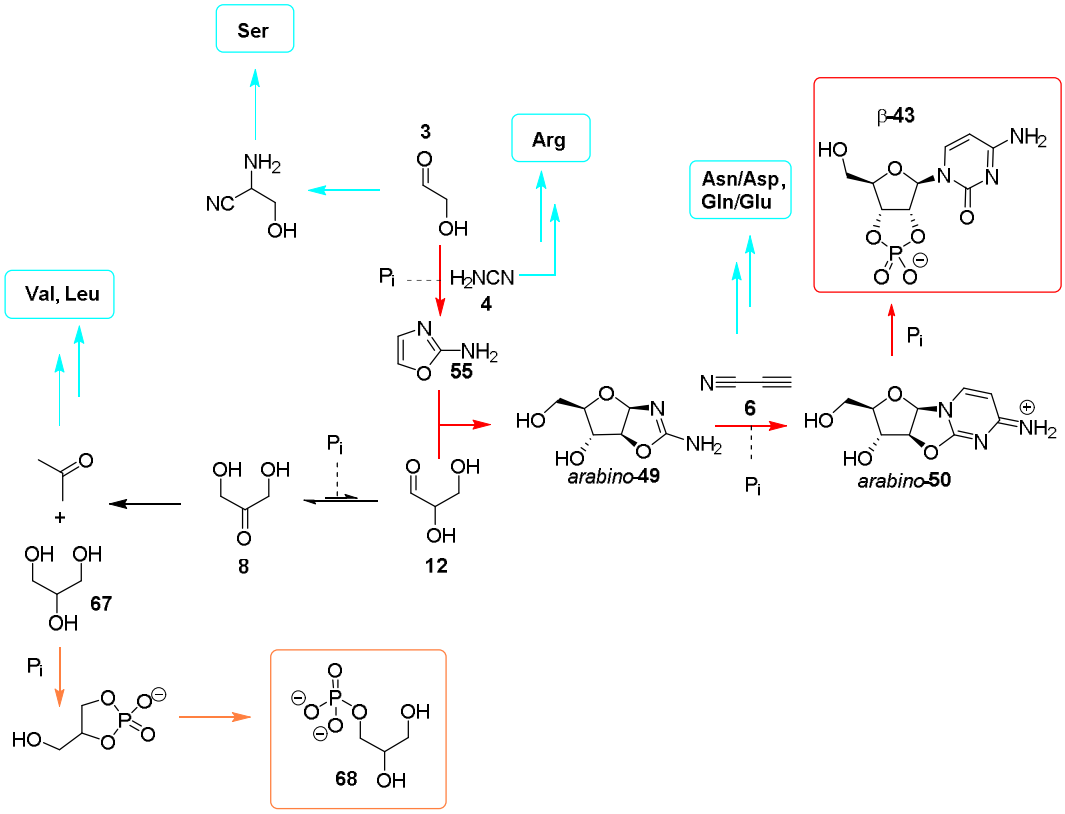


Figure 2.2: Canonical amino acids (three letter codes; blue) and their aldehydic precursors (red), accessed through retrosynthetic analysis as demonstrated on serine (bold box). It is assumed that Strecker synthesis and subsequent hydrolysis on the aldehyde precursors would yield the correct amino acid. Proline isn't included in this retrosynthetic analysis due to its unusual secondary amine backbone structure.

When looking at the aldehydic precursors to amino acids (*Figure 2.2*), we can detect relationships between classes of metabolites that are usually considered distinct. Indeed, both glycolaldehyde (**3**) and formaldehyde (**2**) have been

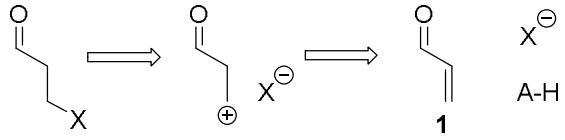
demonstrated experimentally to be precursors to amino acids (serine (**Ser**) and glycine (**Gly**) respectively) through Strecker synthesis[37,276]. Interestingly, both sugars **3** and **2** are also implicated in nucleotide synthesis[33,170,181](*Section 1.2.2*), effectively lying at nodes between these domains. A systems approach to chemistry (*Section 1.3.2*) strives to find compounds like **3** and **2**, that act as generational nodes between different metabolite classes, through cross-referencing of structural and generational relationship of key biomolecules (and their precursors).

In the context of amino acid synthesis, the Sutherland group has developed interconnecting pathways leading to the prebiotic synthesis of 12 precursors to proteinogenic amino acids (of the 20 amino acids found in the genetic code)[37]. Notably, key compounds *en route* to pyrimidine ribonucleotides (as developed by the Sutherland group[33]), can be linked to amino acids, such as: glycolaldehyde (**3**), a precursor to **Ser** (*vide supra*); cyanamide (**4**), involved in the synthesis of arginine (**Arg**); cyanoacetylene (**6**), a precursor to aspartate (**Asp**), asparagine (**Asn**), glutamate (**Glu**) and glutamine (**Gln**, a point to be returned to later); and glyceraldehyde (**12**), a precursor to valine (**Val**) and leucine (**Leu**) through isomerisation of **12** to dihydroxyacetone (**8**) and subsequent reduction to acetone[37] (*Scheme 2.1*). Interestingly, reduction of **8** also produces glycerol (**67**), which can be phosphorylated to phospholipid precursor glycerol-1-phosphate (**68**), thereby linking this reaction network to an additional metabolite class, membranes (*Scheme 2.1*; orange)[37].



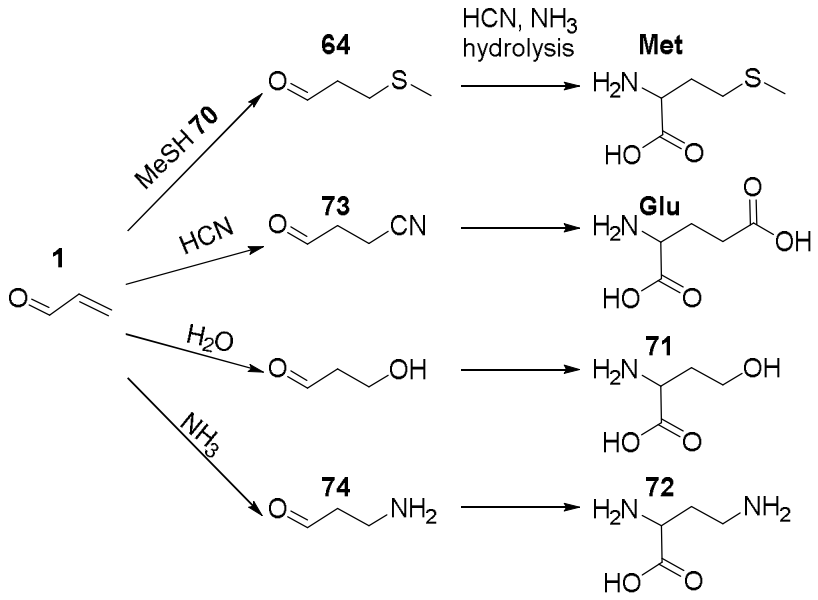
Scheme 2.1: Compounds in the reaction network leading to pyrimidine ribonucleotides (in red) are linked to the synthesis of amino acids (in blue) and membrane precursors (in orange). Adapted from Patel et al.[37] *Steps catalysed by inorganic phosphate (P_i) are shown in dotted lines.*

The *generational* link observed between "distinct" metabolite classes (*vide supra*) led us to explore *constitutional* relationships between metabolites in order to uncover potential generational nodes that have not yet been proposed. We are particularly interested in the constitutional relationship between glyceraldehyde-3-phosphate (**68**), a compound central to metabolism and a potential nucleotide precursor[283], glyceraldehyde (**12**), a known nucleotide precursor[272] (*vide supra*) and amino acids containing a C_3 side-chain. A retrosynthetic analysis of **Glu**, **Gln**, and methionine (**Met**), suggested acrolein (**1**) as a potential synthon (*Scheme 2.2*).



*Scheme 2.2: Retrosynthetic analysis of C_3 aldehydes indicates that these can be obtained from acrolein (**1**), an nucleophile (X) and a general acid (AH).*

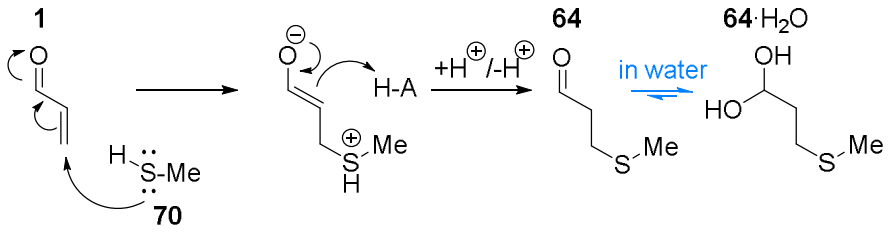
Acrolein (**1**) is the simplest unsaturated aldehyde and it has many prebiotically plausible syntheses, such as aqueous aldol condensation of formaldehyde (**2**) and acetaldehyde (**61**)[284], reaction of methane and water vapour in spark discharge experiments[51] (*vide infra*) or dehydration of glycerol (**67**)[285,286]. There is literature precedent for selective Michael addition to the C3 atom of **1** in Payne's oxidation to form glycidaldehyde (**69**)[287] (a point to be returned to later, *Section 2.2*.). Additionally, Miller and Trump suggested **1** might have been a synthon for the **Met** and **Glu** they observed in their spark-discharge experiments, leading them to study the incubation of **1** (0.8mM), methanethiol (**70**; 0.6 eq), HCN (5 eq) and NH_3 (3.1 eq) at pH 8.7 for 28 days. Following hydrolysis with 3M HCl, they reported synthesis of **Met** (15%) alongside **Glu** (0.5%), and other non proteinogenic amino acids such as homoserine (**71**) and diaminobutyric acid (**72**), that they proposed were formed from reaction of **1** with various nucleophiles followed by *in situ* Strecker reaction and subsequent hydrolysis (*Scheme 2.3*)[51]. This led them to conclude that "acrolein was a key intermediate in prebiotic amino acid synthesis"[51]. We proposed to investigate the extent to which **1** can be used to form some key proteinogenic amino acids (**Glu**/**Gln**, **Met** and **Pro**) selectively in mild, prebiotic conditions, as well as **1**'s role in the prebiotic synthesis of other key metabolites.



*Scheme 2.3: Prebiotic synthesis of amino acids methionine (**Met**), glutamate (**Glu**), homoserine (**71**) and diaminobutyric acid (**72**) from acrolein (**1**) and nucleophiles methanethiol (**70**), HCN, H_2O and NH_3 respectively, followed by Strecker reaction and hydrolysis. Adapted from Trump and Miller*[51].

As mentioned above, the reaction of acrolein (**1**) and methanethiol (**70**) has been suggested as the route to **Met** aldehyde precursor 3-(methylthio)propanal (**64**) in spark-discharge experiments[51] (*Scheme 2.4*). The reaction of **1** and methanethiol (**70**) is also a well-known synthetic process to produce the food-additive **64**[288]. However, synthetic routes to **64** do not seem entirely prebiotically plausible as they are most commonly achieved in neat **1**. We proposed to find mild and prebiotically plausible reaction conditions compatible with our other work, in water and at RT. Experimental difficulties were encountered in the handling of **70** as it has a low solubility in water (23 g/L). Additionally, **70** is volatile, with a boiling point of 6°C, thus reactions done below its pK_a (pK_a 10) lead to loss of **70** as a gas. Therefore, the reactions were conducted by bubbling **70** gas into buffered P_i solution (thereby saturating the solution), while adding a solution of aqueous **1** (pH 7, 0.5M P_i buffer, 140mM **1**, over 5 min). **64** (preferentially sitting as the hydrate **64**·H_2O in a 3:1 ratio) was obtained promptly and near quantitatively (>94% in <30 min, *Figure 2.3, b.*), and was confirmed by spiking with a commercial sample (*Figure 2.3, c.*). The rapidity and selectivity of the reaction

can be explained by the lower solvation of sulphur in water (compared to hydroxyl groups or amines), making methanethiol **70** particularly available as a nucleophile. Furthermore, its "soft" nucleophile nature directs the nucleophilic attack to the β-unsaturated "soft" electrophile moiety of **1**, allowing for a selective synthesis of the desired product **64**.



*Scheme 2.4: Proposed mechanism for the Michael addition of methanethiol (**70**) to acrolein (**1**) forming 3-(methylthio)propanal (**64**) and its hydrate **64**·H_2O.*

Further reaction with cyanide and ammonia (pH 9.5, 1.5 eq KCN, 5 eq. NH_4OH) showed the selective formation of the expected Strecker product 2-amino-4-(methylthio)butanenitrile (**75**)[289] (*Figure 2.3, a.*), further confirming our original interpretation of our acrolein (**1**) and methanethiol (**70**) reaction. Here we show a new efficient way of accessing **64** selectively, in water and in mild prebiotically plausible conditions. Interestingly, due to its lengthy biosynthesis, **Met** has been considered a late addition to the genetic code[24] (*Section 1.1.1*) and therefore an amino acid that was likely not available in the prebiotic mixture on the early Earth. However, we have confirmed that the aldehydic precursor to **Met**, **64**, can be accessed efficiently and selectively from the reaction of two prebiotically available compounds **1** and **70**, leading to the hypothesis that perhaps it was more widely available than was previously thought.

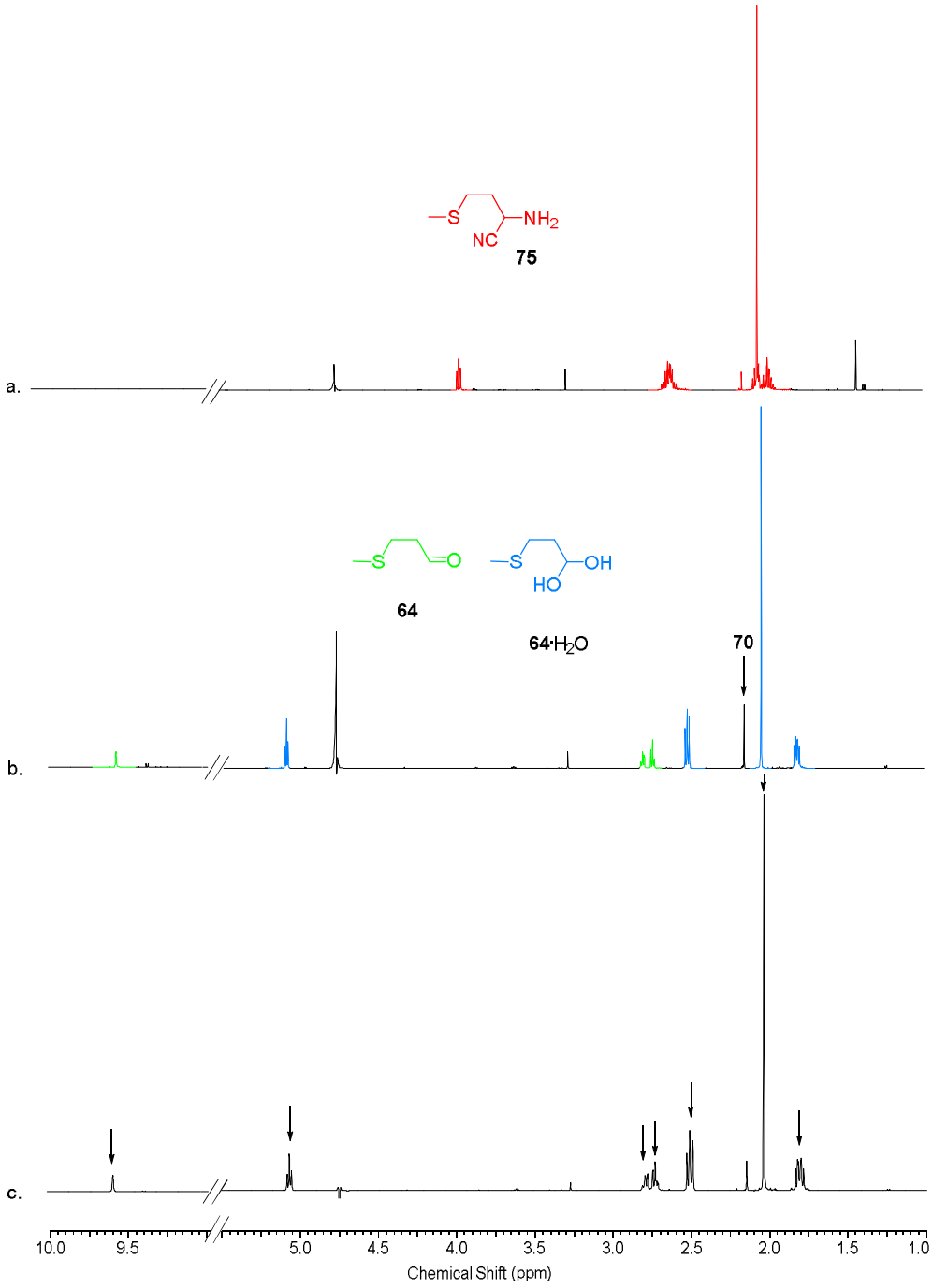
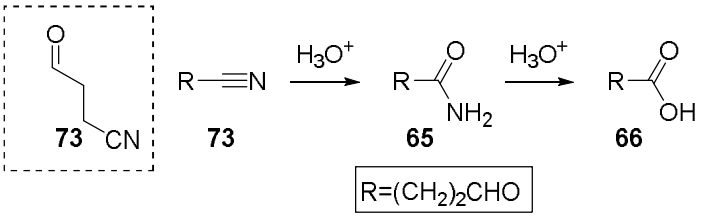


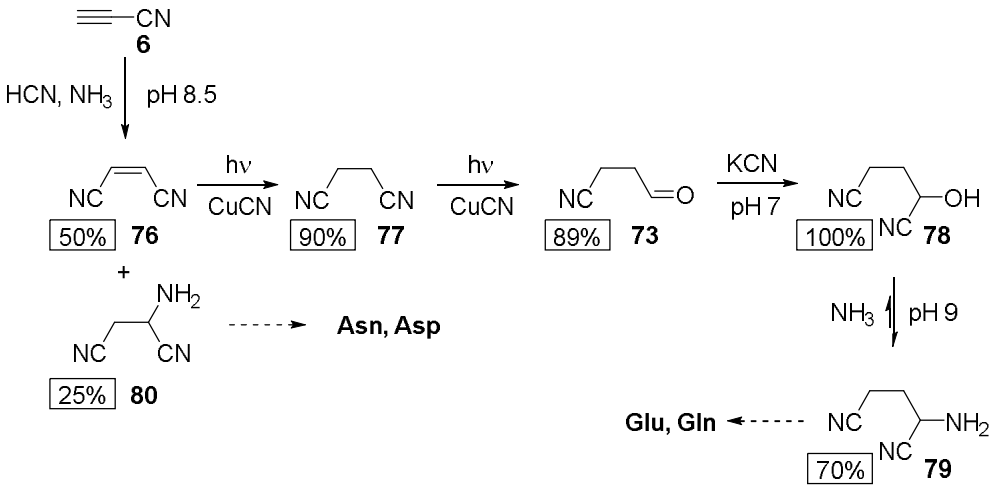
*Figure 2.3: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 10.00-1.00 ppm) a. Strecker synthesis of 2-amino-4-(methylthio)butanenitrile (**75**) from 3-(methylthio)propanal (**64**; 140mM, including hydrate **64**·H_2O; 1.5 eq. KCN, 5 eq. NH_4OH, pH 10, 20 h); b. Reaction of acrolein (**1**; 140mM) with methanethiol (**70**; sat.) at pH 7 after 30 min at RT, showing the synthesis of **64** and its hydrate **64**·H_2O; c. spiked with a synthetic sample of **64**.*

Having established acrolein (**1**) as a precursor to amino acid **Met**, we sought to study its use in the synthesis of other amino acids with a C_3 side chain, such as **Gln** and **Glu**. From the retrosynthetic analysis of amino acids (*vide supra*, *Figure 2.2*), we can see that the precursors for **Gln** and **Glu**, **65** and **66** respectively, are constitutionally similar, with **66** being the hydrolysis product of **65**, which itself can be accessed through hydrolysis of a nitrile 4-oxobutanenitrile (**73**; *Scheme 2.5*).



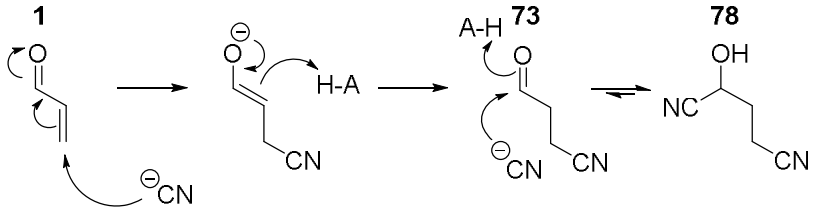
*Scheme 2.5: Aldehydic precursors of glutamine (**65**) can be accessed through hydrolysis of 4-oxobutanenitrile (**73**; dotted box). A second hydrolysis step leads to aldehydic precursors of glutamic acid (**66**).*

Aldehyde **73** has previously been prebiotically synthesised by Patel *et al.*[37] by photoredox chemistry starting with the reaction of cyanoacetylene (**6**) with cyanide to form maleonitrile (**76**) followed by photo-reduction to succinonitrile (**77**) and further photo-reduction to **73** (*Scheme 2.6*).



*Scheme 2.6: Prebiotic synthesis of precursors to amino acids asparagine (**Asn**), aspartic acid (**Asp**), glutamine (**Gln**) and glutamic acid (**Glu**) from cyanoacetylene (**6**), as described by Patel et al*[37].

Trump and Miller studied a one-pot reaction of acrolein (**1**; 0.8mM), HCN (5 eq) and NH_3 (3.1 eq) at pH 8.7 for 28 days, followed by hydrolysis with 3M HCl yielding **Glu** (1.5%), supposedly through *in situ* synthesis of 4-oxobutanenitrile (**73**) and subsequent Strecker reaction and hydrolysis[51]. We proposed to revisit the reaction of **1** and cyanide in mild, aqueous conditions, in the hopes of accessing **73** through Michael addition of cyanide to **1** (*Scheme 2.7*). The high selectivity and reactivity of HCN addition to aldehydes, forming cyanohydrins[282], leads us to expect the formation of cyanohydrin 2-hydroxypentanedinitrile (**78**) selectively upon reaction with an excess of HCN (*Scheme 2.7*).



*Scheme 2.7: Proposed mechanism for the Michael addition of cyanide to acrolein (**1**) forming 4-oxobutanenitrile (**73**) and its cyanohydrin 2-hydroxypentanedinitrile (**78**).*

As predicted, we obtained full conversion to cyanohydrin **78** through reaction of **1** (140mM) with an excess of KCN (700mM; pH 9). ^{1}H NMR analysis (*Figure 2.4*, *a.*) showed the characteristic peaks for H_2-C2 at 2.68 ppm (t, J = 7.1 Hz) and H_2-C3 at 2.17 ppm (complex multiplet) of **78**, in accordance with literature data for **78**[37] (H_2-C2: 2.62 ppm, t, J = 7.2 Hz and H_2-C2: 2.15 ppm, m). Further reaction with NH_3 (5 eq, pH 10.5) showed the formation of the expected Strecker product, 2-aminopentanedinitrile (**79**)[37,289] (*Figure 2.4*, *b.*) further confirming our original interpretation of **78** formation from our **1** and HCN reaction. Here we show a new way of accessing precursors to **Gln** and **Glu** selectively in water and mild conditions, avoiding the need for photoredox chemistry, and the lengthy synthesis described by Patel *et al.*[37]. Furthermore, the synthesis of amino acid precursor **78** described here confirmed acrolein (**1**) as a precursor for amino acids **Gln** and **Glu** and confirmed our hypothesis that **1** can be used as synthon for multiple amino acids with a C_3 side chain by simple Michael addition in water.

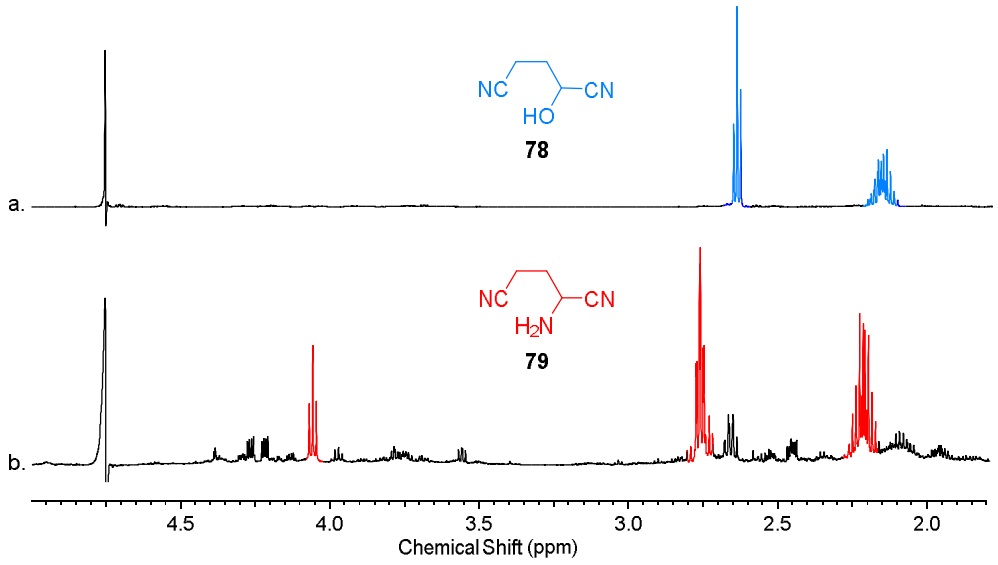
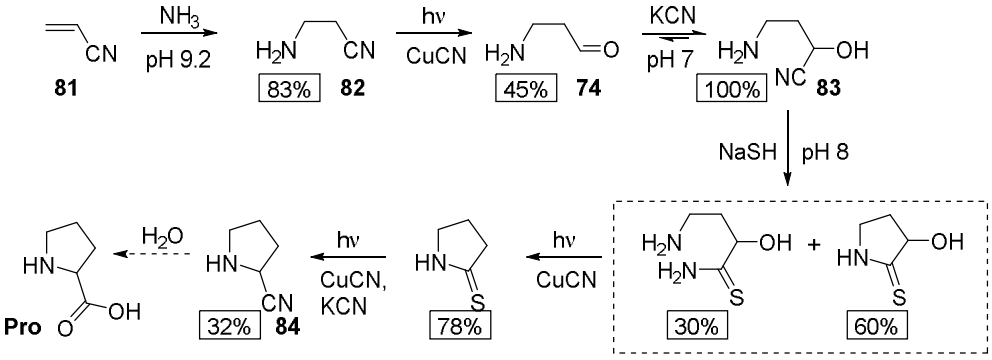


*Figure 2.4: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 5.00-1.75 ppm) a. Reaction of acrolein (**1**; 140mM) with KCN (5 eq) at pH 9 after 4 h at RT, showing the synthesis of 2-hydroxypentanedinitrile (**78**; blue) – the H-(C1) signal is lost under the residual solvent peak; b. Strecker product 2-aminopentanedinitrile (**79**; red) from incubation of **78** (140mM) with ammonia (5 eq) at pH 10.5 after 8 h at RT.*

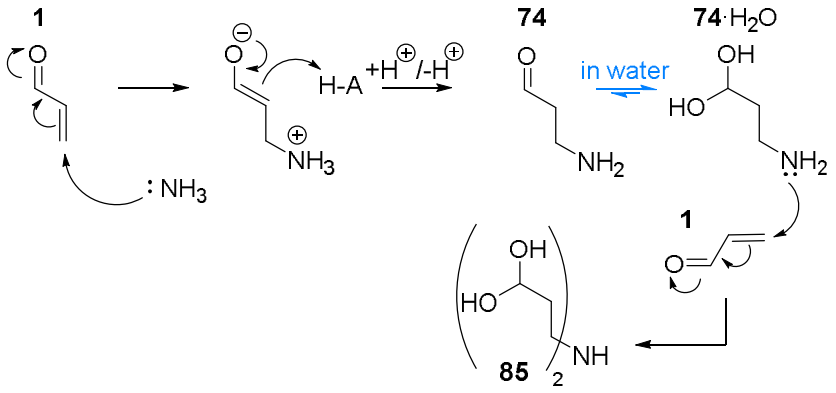
Another amino acid that contains a C_3 unit is proline (**Pro**). **Pro** is unusual among amino acids as it is the only proteinogenic amino acid with a side-chain cyclised onto its backbone nitrogen, forming a secondary amine. Because of this structural difference, Patel *et al.* proposed an alternative prebiotic route to **Pro** that does not involve the traditional Strecker reaction[37] (*Scheme 2.8*). Instead, they devised a 7 step synthesis starting with the reaction of ammonia and acrylonitrile (**81**) to form aminopropionitrile (**82**), followed by photoreduction to 3-aminopropanal (**74**). Once aldehyde **74** is accessed, they form the cyanohydrin **83**, followed by reaction with hydrosulphide and photoredox chemistry to access the aminonitrile precursor of **Pro**, **84**.



*Scheme 2.8: Prebiotic synthesis of precursors to amino acid proline (**Pro**), from acrylonitrile (**81**), as described by Patel et al*[37].

Although the pathway to **Pro** devised by Patel *et al.*[37] is synthetically distinct from other prebiotic amino acid syntheses described previously, it does incorporate many of the same reactants, such as ammonia and cyanide. Interestingly, the synthesis described above occurs *via* the C_3 aldehyde 3-aminopropanal (**74**). The one-pot reaction of acrolein (**1**; 0.8mM), HCN (5 eq) and NH_3 (3.1 eq; pH 8.7 for 28 days) designed by Trump and Miller[51] (*vide supra*), yielded the amino acid derived from aldehyde **74**: diaminobutyric acid (**72**; 0.8% after hydrolysis with 3M HCl), alongside **Glu**. They proposed that this amino acid was formed through *in situ* reaction of **1** with ammonia to form **74** and subsequent Strecker reaction and hydrolysis[51] (*Scheme 2.3*, *vide supra*).

We sought to improve the yield and selectivity of **74** synthesis through Michael addition of NH_3 to **1** in water (*Scheme 2.9*). Reaction of **1** (140mM) with an excess of NH_3 (700mM, pH 6.5) showed, through ^1H NMR analysis, formation of two major species: **74** and dimer **85** (*Figure 2.5, a.*).



*Scheme 2.9: Proposed mechanism for the Michael addition of ammonia to acrolein (**1**) forming 3-aminopropanal (**74**) and its hydrate **74**·H_2O. Proposed mechanism for the formation of dimer **85**.*

Comparing the crude reaction spectrum with data obtained from synthetically prepared **74** (*via* deprotection of 1-amino-3,3-diethoxypropane **86**, *Figure 2.5*) and comparing to previously reported ^1H NMR data for **74**[37] confirmed our interpretation of one of the major products as **74**. Due to the similarity in coupling between **74** and the second major species present **85** (**74**·H_2O - H-C1: 5.21 ppm, t, J = 5.4 Hz; H_2-C3: 3.12 ppm, t, J = 7.0 Hz; and H_2-C2: 1.95 ppm, dt, J = 7.0, 5.4 Hz *vs* **85** - H-C1: 5.16 ppm, t, J = 5.6 Hz; H_2-C3: 3.63 ppm, t, J = 6.8 Hz; and H_2-C3:1.84 ppm, dt, J = 6.8, 5.6 Hz), and the increased reactivity of primary amines compared to ammonia (due to the electron donating alkyl group), we proposed **85** is most likely a dimer[iv] formed from the Michael addition of **74** to starting material **1**.

[iv] No further studies have been carried out to ascertain the possible structures.

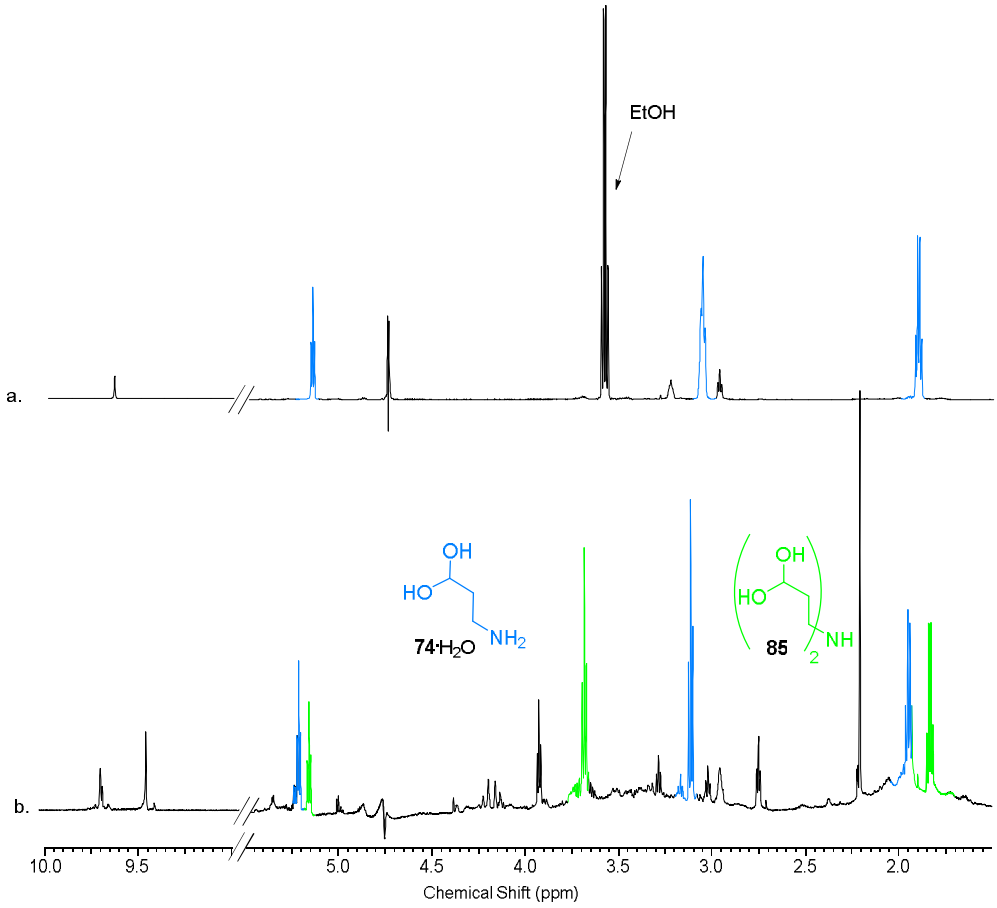


*Figure 2.5: ^1H NMR spectra (600 MHz, H_2O/D_2O 9:1, 10.00-1.50 ppm) a. synthetic preparation of 3-aminopropanal (**74**; seen predominantly in its hydrate form **74**·H_2O) from deprotection of 1-amino-3,3-diethoxypropane (**86**); b. Reaction of acrolein (**1**; 140mM) with NH_3 (5 eq) at pH 6.5 after 12 h at RT, showing the synthesis of 3-aminopropanal hydrate (**74**·H_2O; blue) and another species proposed to be the 3-aminopropanal dimer **85** (green).*

Pro is unique among natural amino acids as it is the only proteinogenic amino acid with a secondary amine due to the cyclic nature of the side-chain onto the backbone. This implies it has to be accessed *via* a different prebiotic pathway than other amino acids, such as the one described by the Sutherland group (*Scheme 2.8*), as Strecker synthesis would lead to non-natural amino acid **72**. We studied the reaction of the aldehyde **74** we obtained with cyanide in the hopes of obtaining the cyanohydrin 4-amino-2-hydroxybutanenitrile (**83**). The excess NH_3 still present from the initial acrolein (**1**) and ammonia reaction meant that the reaction needed to be carried out at low pH in order to limit Strecker reaction. Indeed, the

pH dependence of aminonitrile *vs* cyanohydrin synthesis has been studied extensively and shows that an aldehyde, in the presence of ammonia and cyanide, will sit near exclusively as the cyanohydrin when the reaction is at pH<6 and near exclusively as the aminonitrile when the pH>9[282]. This is due to ammonia being pulled out of the equilibrium as its ammonium salt at low pH (pK_a 9.25). The reaction was therefore carried out at low pH (pH 3, 1.5 eq. KCN), which lead to selective formation of the cyanohydrin 4-amino-2-hydroxybutanenitrile (**83**; 6 h at RT, *Figure 2.6, b.*). A similar reaction done with the synthetically prepared **74** shows the formation of the same species **83** (*Figure 2.6, a.*) confirming our interpretation of cyanohydrin **83** formation. Through these studies we showed a new way of accessing **83** in a one-pot reaction in water and in conditions compatible with the synthesis of precursors to amino acids **Met** and **Glu**/**Gln**.

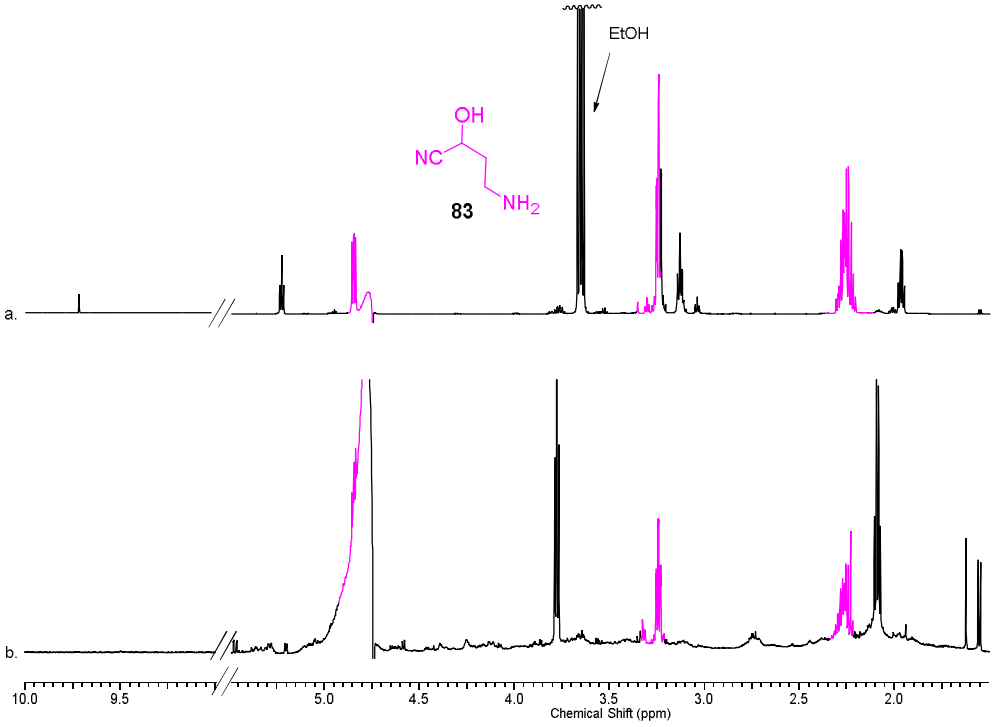
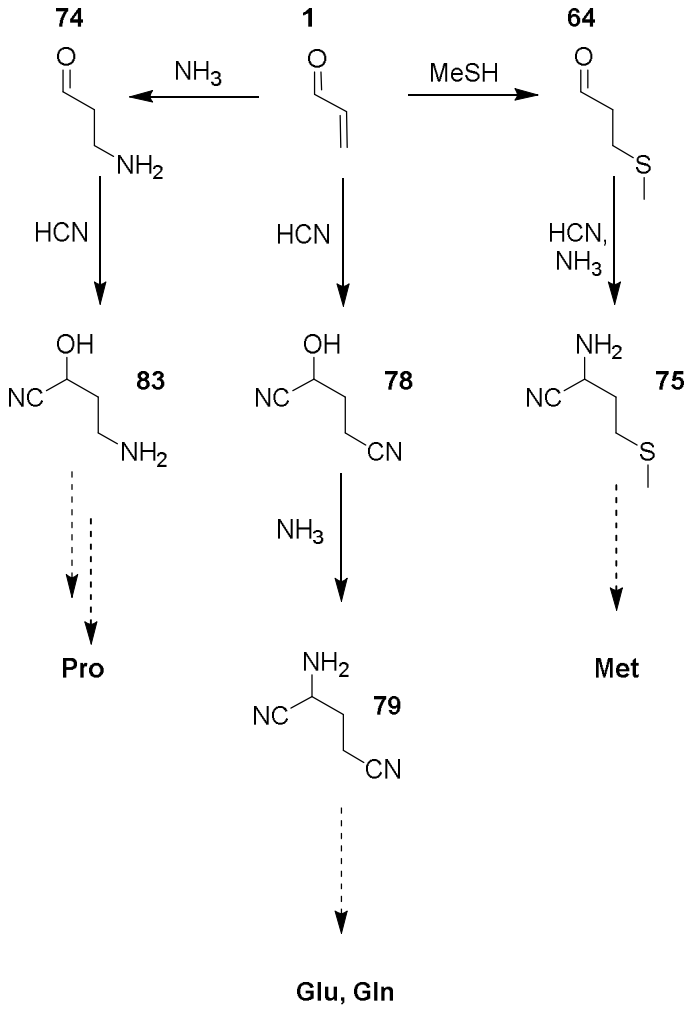


*Figure 2.6: ^1H NMR spectra (600 MHz, H_2O/D_2O 9:1, 10.00-1.50 ppm) a. synthetic preparation of 3-aminopropanal hydrate (**74**·H_2O) incubated with KCN (1:1, pH 3, 6 h); b. Reaction of prebiotically prepared **74**·H_2O with KCN (1.5 eq), at pH 3 after 6 h at RT showing the synthesis of 4-amino-2-hydroxybutanenitrile (**83**; pink) – the H-(C1) peak (dd) can be seen partially supressed by the residual solvent peak.*

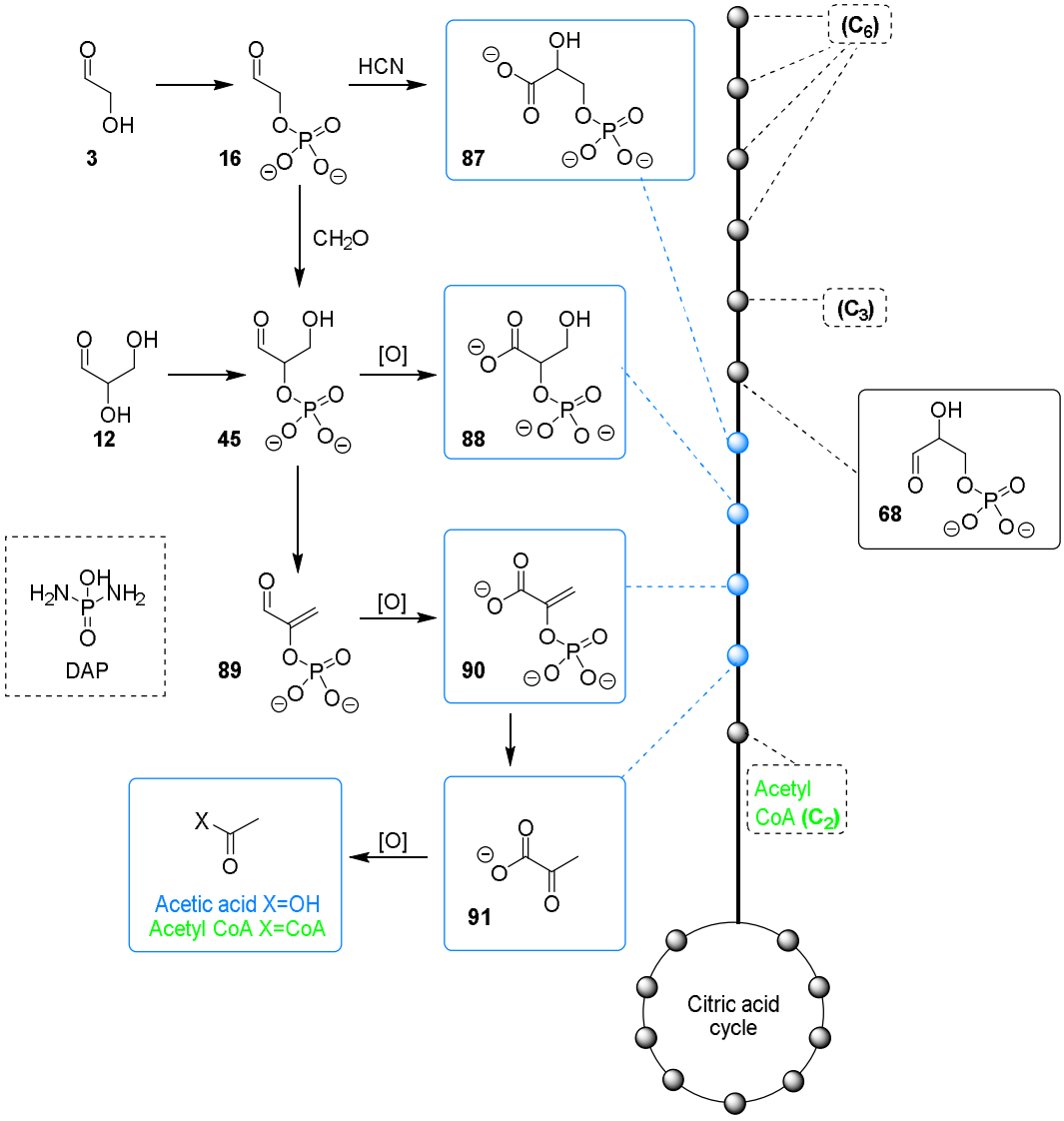
These syntheses described above solidify acrolein (**1**) as a central C_3 prebiotic feedstock molecule that can be used in water to efficiently form aldehydic precursors to some of the proteinogenic amino acids (**Pro, Glu**/**Gln** and **Met**; *Scheme 2.10*). Due to the constitutional simplicity of **1**, we proposed to explore other chemistries in avenues relevant to prebiotic synthesis, such as nucleotide synthesis.



*Scheme 2.10: Prebiotic synthesis of aldehydic precursors to **Pro**, **Glu**/**Gln** and **Met** by Michael addition of NH_3, HCN and MeSH to acrolein (**1**), respectively.*

2.2. Oxidation of acrolein

Even though the obvious interconnectedness of amino acids and nucleic acids in biology cannot be disputed (as can be seen in the conversation of both in the genetic code[12,20]), suggesting a common origin[32–37], the problem of their prebiotic origin has traditionally been studied separately. This is because of the assumption that the oxygenous chemistry necessary for the synthesis of sugars (such as ribose (**9**), through the formose reaction) is not compatible with the nitrogenous chemistry necessary for the synthesis of bases and amino acids[190] (*vide supra*, *Section 1.3.2*). However, this separation effectively reduces the reaction space and greatly limits the potential of new interconnected chemistries being found. Recent advances have been made by looking at how all the different parts of biology, and the chemistry that led to it, connect[37,276]. This systems chemistry approach, such as the study of prebiotic amino acid and nucleotide synthesis mentioned above (*Section 2.1.*) shows new connections between previously disconnected metabolite classes. We previously described how compounds implicated in nucleotide synthesis, such as glycolaldehyde (**3**) and glyceraldehyde (**12**), can be linked to amino acids[37] (*vide supra*, *Scheme 2.1*). Recent work in the Powner lab shows that both **3** and **12** can also be linked to metabolism through phosphorylation and oxidation[290] (*Scheme 2.11*). Coggins and Powner describe a prebiotic network, starting with **3** and **12** that "mimics a metabolic trajectory with absolute specificity under mild aqueous conditions from simple prebiotic aldehydes".



*Scheme 2.11: Prebiotic link between nucleotide precursors glycolaldehyde (**3**) and glyceraldehyde (**12**) and the highly conserved triose glycolysis metabolic network (blue) leading to the citric acid cycle through phosphorylation and oxidation, as reported by Coggins and Powner*[290].

They found that by incubating **3** or **12** (25mM) with diamidophosphate (DAP) (100mM) they could access the α-phosphorylated products glycolaldehyde-phosphate (**16**) or glyceraldehyde 2-phosphate (**45**) selectively and efficiently (90-96% yield after 2 h at pH 4 and RT)[290]. Incubation of **16** (200mM) with HCN (in P_i buffer, 750mM at pH 12, 75°C) yielded key metabolite glyceric acid 3-phosphate (**87**) through transient formation of the cyanohydrin followed by

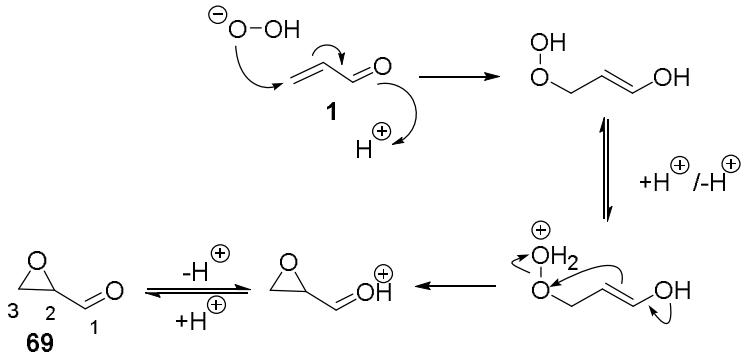
hydrolysis[290]. Sugar-phosphate **45** can yield two different metabolites through varied incubation conditions. Oxidation of **45** (350mM) with sodium chlorite (1.4 eq in P_i buffer 60mM, pH 4-7) yields glyceric acid 2-phosphate (**88**) quantitatively; whereas heating **45** (80mM) in P_i buffer (0.5M) at 60°C for a day leads to a divergent path of efficient dehydration to phosphoenol pyruvaldehyde (**89**; 74%)[290]. This high-energy phosphoenol ether **89** can then lead directly, through oxidation, to the highest energy phosphate used in biology: phosphoenol pyruvate (**90**)[290,291] (*Scheme 2.11*, *vide supra*). Coggins and Powner describe four different oxidative strategies used (while detailing the prebiotic feasibility of each reactant used): cyanohydrin formation coupled with ferricyanide; manganese dioxide coupled with cyanide; Fenton oxidation; and chlorates. The latter is the most efficient with quantitative conversion being observed after incubation of **89** (70mM) with sodium chlorite (1.4 eq in 60mM P_i buffer, pH 4-7), in the presence of a hypochlorite scavenger such as methionine (**Met**). They conclude that the "facile oxidation of [**89**] in water is extremely general and thus prebiotically plausible"[290].

Coggins and Powner complete the prebiotic network by demonstrating that heating phosphoenol pyruvate (**90**, 63mM; 60°C, pH 4-7) led to quantitative conversion to pyruvate (**91**) after 2 days. **91** feeds directly into the highly conserved citric acid cycle. Lastly, they found that oxidation and decarboxylation of **91** gives access to acetate, which composes the carbon framework found in essential co-enzyme acetyl-coA, leading Coggins and Powner to remark that "The direct liberation of [acetate] by oxidation of [**91**] suggests an essential prebiotic link between C_3 and C_2 metabolism within our system"[290].

The high chemical control observed in Coggins and Powner's oxidative network suggests the role of oxidation in prebiotic chemistry warrants renewed attention. Oxidation of acrolein (**1**) to form glycidaldehyde (**69**) has been reported by Payne as a selective and efficient reaction[287]. The constitutional simplicity and robust prebiotic synthesis of **1**[51,284–286], coupled to its propensity to form proteinogenic amino acid precursors by Michael addition (*vide supra*, *Section 2.1*), lead us to

study its place in ribonucleotide synthesis, in the hopes of establishing **1** as a generational node between different metabolite classes.

We sought to access epoxide **69** through oxidation of **1** with hydrogen peroxide (H_2O_2; *Scheme 2.12*). H_2O_2 is a constitutionally simple, prebiotically robust oxidant that has many prebiotically plausible syntheses, including photolysis of water vapour in the atmosphere that can then be "scavenged efficiently by cloud droplets and by rain"[292]. Borda *et al.* also demonstrated that pyrite-induced hydroxyl radical formation in water leads to H_2O_2 formation[293,294]. H_2O_2 can also form at low temperature[53] both in the reaction pathway of surface hydrogenation of O_2 that could occur in the interstellar medium[295], and in hydrothermal vent conditions[296]. Oxidation of **1** (140mM) with H_2O_2 (1 eq) through modification of Payne's literature protocol[287] (pH 8.5) afforded a near quantitative conversion (>90%) (*Figure 2.7*) and a 27% isolated yield of **69** (through extraction and subsequent distillation).



*Scheme 2.12: Mechanism for oxidation of acrolein (**1**) with hydrogen peroxide to form glycidaldehyde (**69**).*

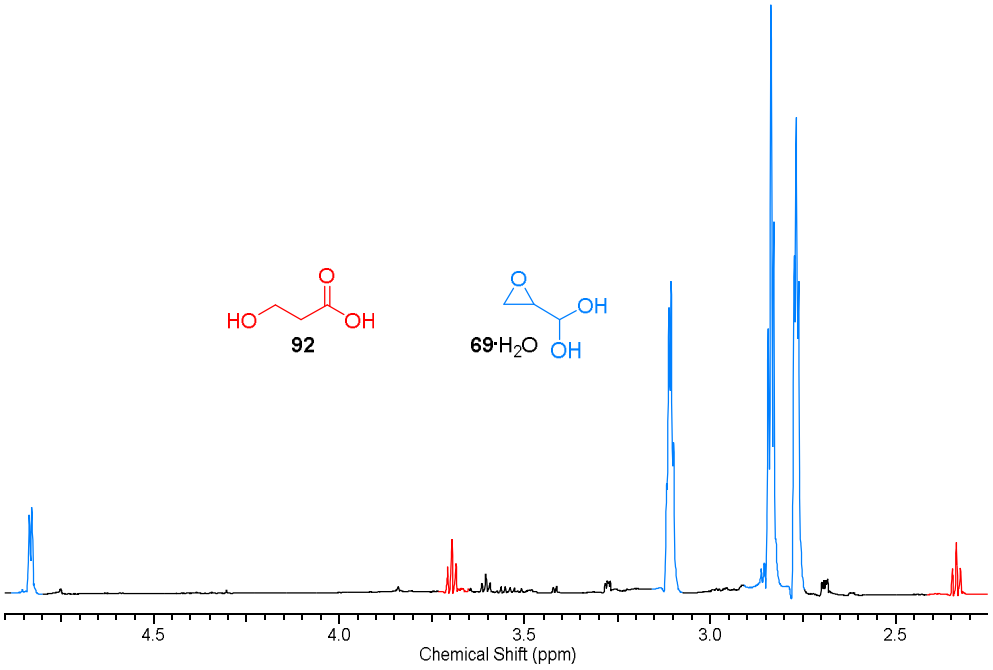
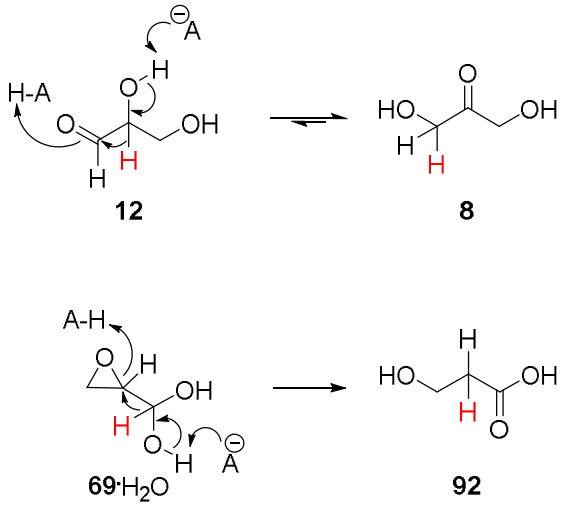


*Figure 2.7: ^{1}H NMR spectrum (600 MHz, H_2O/D_2O 9:1, 4.90-2.00 ppm) of the reaction of acrolein (**1**; 140mM) with H_2O_2 (1 eq) at pH 8.5 after 2 h at RT, showing the synthesis of glycidaldehyde (**69**), sitting as the hydrate form **69**·H_2O (minor amount of 3-hydroxypropanoic acid (**92**), 8%, was observed).*

No aldehyde signal was observed for **69** suggesting it behaves similarly to glycolaldehyde (**3**) and glyceraldehyde (**12**) (and ribose (**9**) when in its open-chain form[209]) and sits as the hydrated form **69**·H_2O. We propose that observed by-product 3-hydroxypropanoic acid (**92**) was formed through hydride shift (*Scheme 2.13*, bottom) of **69**·H_2O, in a similar manner to one of the mechanistic routes for the aldose/ketose isomerisation of glyceraldehyde (**12**) to dihydroxyacetone (**8**)[297,298] (*Scheme 2.13*, top).



*Scheme 2.13: Mechanism for the 1,2-hydride shift in the aldose/ketose isomerisation of glyceraldehyde (**12**) to dihydroxyacetone (**8**) as demonstrated by Nagorski and Richard[297], and Appayee and Breslow[298] (top). Proposed mechanism for the formation of 3-hydroxypropanoic acid (**92**) from glycidaldehyde hydrate **69**·H_2O (bottom).*

This simple and highly prebiotically plausible protocol allowed rapid access to **69** at a scale and purity sufficient to explore its reactivity. A study of its hydrolysis over time was performed in order to better assess its background reactivity in the solvent used. Therefore, we undertook an investigation of the pH dependent hydrolysis rate of **69** in water, in which we observed generation of glyceraldehyde (**12**; *Figure 2.8*). The hydrolysis product **12** was confirmed by spiking with an authentic sample (Appendix – *Figure 11.7*). These results suggest a direct way of generating a link between acrolein (**1**) and a sugar implicated in ribonucleotide synthesis[33,272] through the hydrolysis of **69** to **12**.

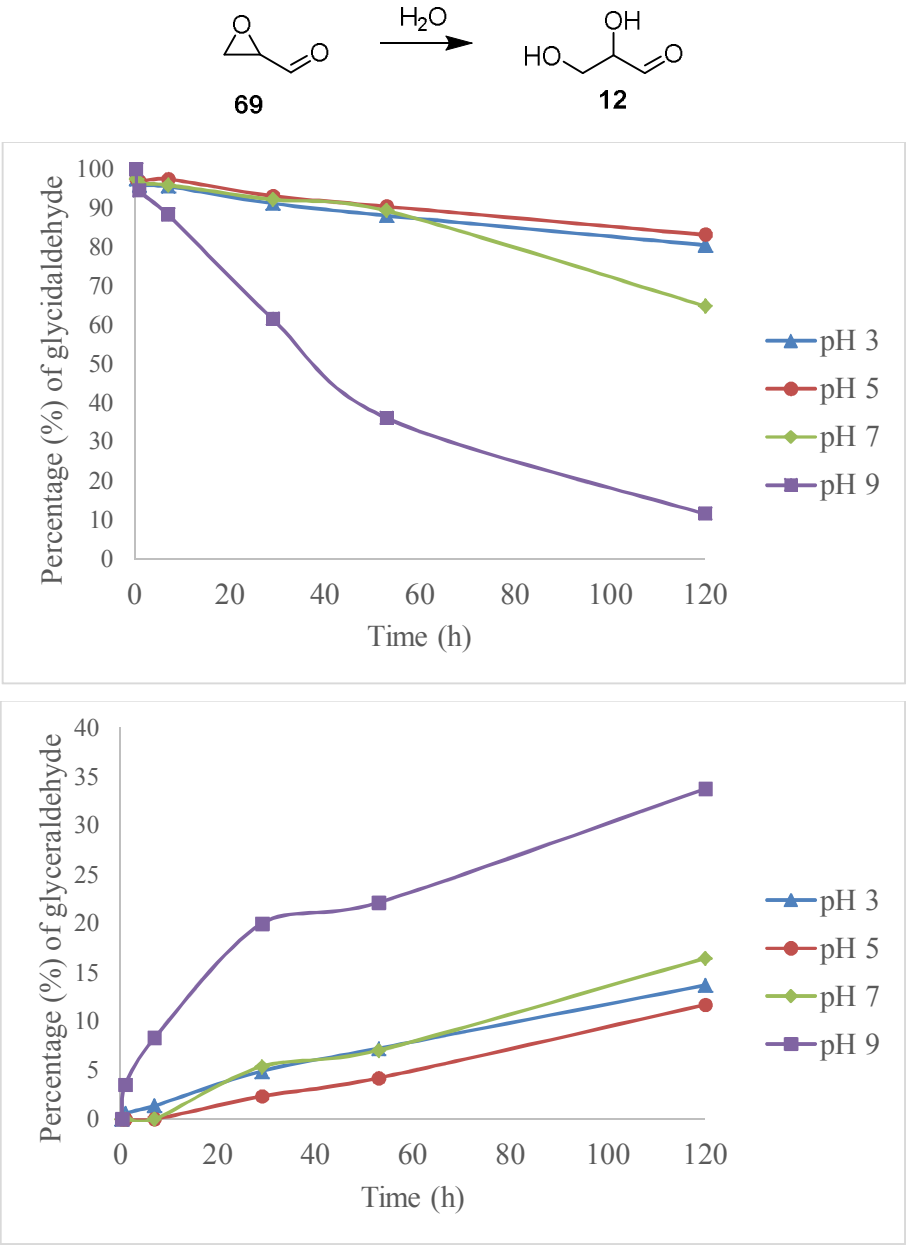


*Figure 2.8: Comparison of hydrolysis rates of glycidaldehyde (**69**; 400mM, RT) at different pHs over time (percentage is calculated based on internal standard). Top: Degradation of glycidaldehyde (**69**) over time. Bottom: Formation of glyceraldehyde (**12**) over time.*

Interestingly, glycidaldehyde (**69**) showed a greater stability at lower pH. The half-life of **69** (400mM) in water at pH 9 and RT was calculated to be 40 h, whereas it is observed to be of the order of 230 h for pH 7 (extrapolation from the observed hydrolysis at 8 days) at RT. These initial results suggest that there is a

window of stability, at neutral pH, where a low hydrolysis rate of **69** may afford an opportunity to retain the novel reactivity of **69** compared to **12**.

We have developed a new hypothesis that this epoxide product, glycidaldehyde (**69**), may have been a central compound in prebiotic ribonucleotide synthesis due to its constitutional relationship with glyceraldehyde (**12**). Specifically, we envisaged that it could replace **12** as a triose source while having the additional benefit of being more reactive by virtue of dehydration. Specifically, the epoxide moiety opens the potential for novel reactivity not accessible with **12**, such as nucleophilic insertion at C3 (*vide infra*, *Section 3*). Even though **12** is considered a precursor to ribonucleotides[37,283], it suffers from several draw backs such as its propensity to isomerise to its more stable form dihydroxyacetone (**8**)[297,298], especially when in the presence of a general acid/base catalyst such as phosphate, often used in prebiotic synthesis for its buffering capacities as well as its role in the constitution of ribonucleotides[297]. Isomerisation to **8** makes **12** unavailable for reaction, and in conditions relevant to nucleotide synthesis this isomerisation to the ketose **8** leads to synthesis of non-natural branched nucleotides (a point to be returned to later, *Section 5*). We hypothesise that **69** would solve these problems due to its inability to react through its masked hydroxyl groups.

Through this oxidative strategy, and the study of the Michael addition of various nucleophiles to acrolein (**1**), we have shown that **1** is a synthon in the generation of both amino acids and sugars (potential ribonucleotide precursors). Finding generational nodes such as this helps bridge the gap between previously separate parts of prebiotic chemistry research and helps bring us closer to a unified prebiotic synthesis of all of life's chemistry, the ultimate goal in prebiotic chemistry.

3. Prebiotic incorporation of phosphate

3.1. Previous attempts at prebiotic phosphorylation

The incorporation of phosphate into nucleosides or its precursors is necessary for a complete abiogenesis of nucleotides[214] (*Section 1.2.5*). However, this has proved to be a notoriously difficult step to achieve in prebiotic conditions[223,225–228]. The historic use of phosphate as an electrophile in prebiotic chemistry has been a barrier to aqueous phosphorylation, as the competition of the solvent -water- with any other alcohol that one might want to phosphorylate greatly limits its potential use[223]. Attempts to enhance the electrophilic nature of phosphate have been sought, such as the use of condensing agents (such as cyanamide (**4**), cyanoformamide (**37**), cyanogen (**5**), carbodiimide (**93**), *Figure 3.1*), but again the reactions are inefficient due to the great excess in concentration of water (55M), leading to low yields of phosphorylation of nucleotides (<4%)[223].

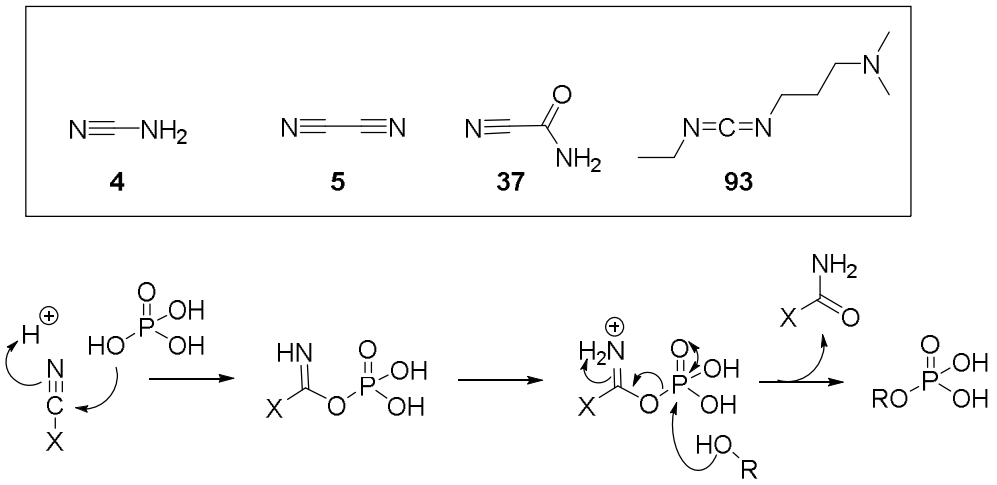
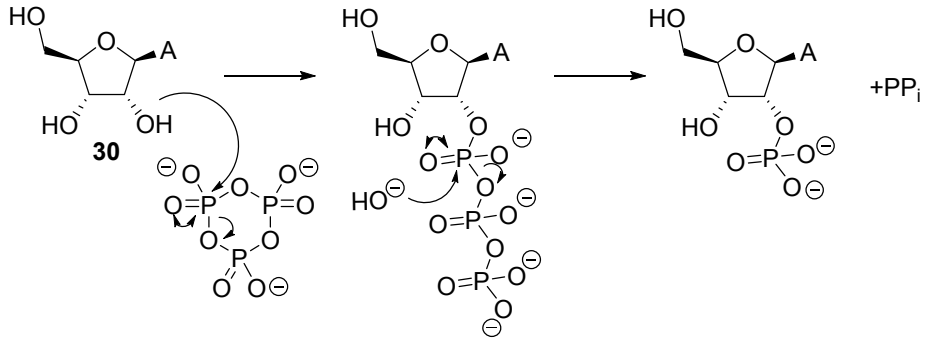


Figure 3.1: Use of condensing agents (box) in the activation of phosphate.

Cyclic trimetaphosphate in strongly alkaline conditions leads to high yielding C2′ and C3′ specific phosphorylation of nucleotides (*Scheme 3.1*), with the highest yield, 90%, obtained on riboadenosine (**30**; 0.1M) after 4 days in 1M NaOH, with 10 eq tri(tetramethylammonium) salt of cyclic trimetaphosphate, at RT. However, the high pH of the reaction and questionable prebiotic availability of cyclic

trimetaphosphates has put into question the prebiotic relevance of the reaction[63,225].



*Scheme 3.1: Proposed mechanism for the aqueous phosphorylation of nucleosides as reported by Saffhill[224], as demonstrated on the 2′ hydroxyl of riboadenosine (**30**; where A corresponds to adenine). The phosphorylation also occurs on the 3′-hydroxyl. PP_i corresponds to pyrophosphate.*

In order to limit the competition of water, non-aqueous solvents and dry-state phosphorylations of nucleosides have been studied extensively[225–228]. Non-aqueous solvent phosphorylation studies by Schoffstall showed good yields of nucleoside phosphorylation in pure formamide (**36**; 50% phosphorylated products after 15 days, 70°C)[226]. However the same reaction in wet **36** (20% water, 80% **36**) led to dramatic decrease in yields (2% mono-phosphate compounds formed)[226]. Beck *et al*. have shown that dry-heating uridine with inorganic phosphate at 65°C for 9 months can also produce about 50% yield of nucleotides[227]. Further studies by Lohrmann and Orgel have shown improved efficient dry-state phosphorylation (<90%) using ammonium phosphate, in the presence of urea (**28**), in only a few hours, and more moderate yields when using a prebiotic mineral phosphate source, hydroxylapatite (20% yield phosphorylated compounds after 24 h)[225]. Despite the efficiency of these non-aqueous reactions, the problem remains one of selectivity, as all these reactions yield a mixture of mono-, di- or cyclic phosphates[225–228] (see *Figure 1.13*, page 55).

In 2009, Powner *et al*. showed dramatic improvement on selectivity of the dry-state phosphorylation of nucleosides by submitting furanosyl-2,2′-anhydrocytidines (**50**) to urea-inorganic phosphate, or urea-pyrophosphate, dry-

state phosphorylations[33]. The structure of **50** is proposed to affect the susceptibility of the different hydroxyl groups with regards to phosphorylation. Choudhary *et al*. showed that the n→π* interaction of O5′ and C2 orbitals in *arabino*-**50** leads to a decrease in the nucleophilicity of O5′ and thus making O3′ comparatively more reactive and available for phosphorylation[299] (*Figure 3.2*).

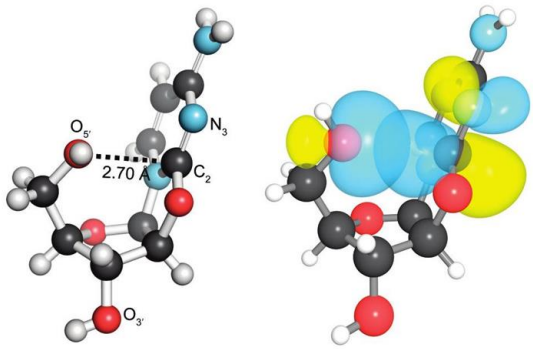


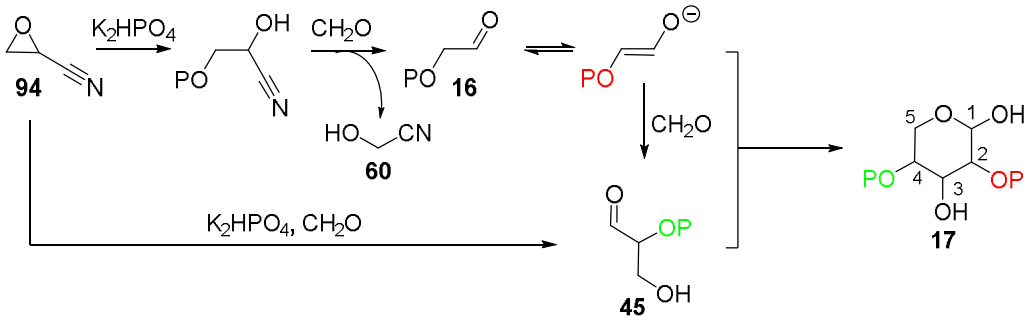
*Figure 3.2: Ball-and-stick representation of D-arabinofuranosyl-2,2'-anhydrocytidine (arabino-**50**) showing interaction between O5' and C2 orbitals, leaving O3' prone to phosphorylation. Left: From crystal structure; Right: From computational study. Adapted from Choudhary et al*[299].

Even though significant improvements have been made, one has to question the prebiotic feasibility of non-aqueous reactions. Dry-state heating can be imagined in a number of prebiotic situations such as "in desert areas where temperatures in excess of 70°C are encountered"[225]. However, considering the majority of prebiotic reactions leading to nucleosides, or its precursors, are described in aqueous solutions, requiring dry-state heating mid-synthesis would limit the abundance of possible environments. Similarly, there is a case for using formamide (**36**) as a solvent, as defended by Schoffstall[226]. **36** is a well-accepted prebiotic molecule[226,300] that is less-volatile than water, which has led to suggestions that heating pools of concentrated **36** in water could lead to incremental increases in **36** concentration. However, based on the abundance of water on Earth (and on the early Earth[62]), one might argue that pure pools of **36** would have been rare. If any pools of **36** could form, one would expect some level hydration, which would lead to significant decreases in phosphorylation yields, as

observed by Schoffstall (*vide supra*)[226]. In conclusion, a way of selectively and efficiently phosphorylating nucleosides, or its precursors, in water remains to be found, in order to increase the prebiotic likely-hood of the reaction, while staying consistent with the environment used in other prebiotic reactions.

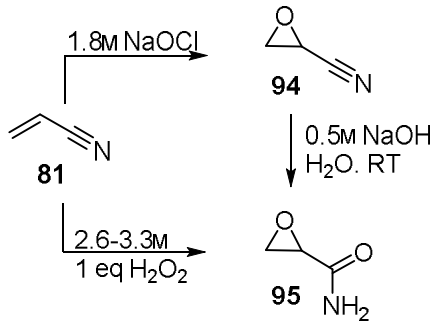
3.2. Nucleophilic aqueous phosphorylation

Naturally occurring nucleotides are linked by a 3′-5′ phosphodiester bond, therefore a prebiotic route to C5′ or C3′ phosphorylated nucleotides is essential to find (*vide supra*, *Section 1.2.5*). The Eschenmoser group have previously reported nucleophilic aqueous phosphorylation of oxiranecarbonitrile (**94**) during their seminal studies of sugar-phosphate aldol chemistry[301] (*Scheme 3.2*). After an initial reaction of phosphate (10 eq), the nitrile moiety can be lost by addition of formaldehyde (**2**, 1.2 eq; to form glycolonitrile (**60**)), yielding glycolaldehyde phosphate **16** (47% yield, after 25 h at pH 10.5). Prolonged incubation (for 13 days) with formaldehyde (**2**; 20 eq) yields glyceraldehyde-2-phosphate (**45**) *via* aldol reaction (30% yield, 10 eq K_2HPO_4, at pH 10.5). Combining the two phosphorylated species obtained, **45** and **16** (37mM), can lead to efficient conversion to pentose-2,4-diphosphates (60% after 14 days in 2M NaOH) with a preference for the ribose-2,4-diphosphate (**17**; 43% of the pentose-2,4-diphosphates[273], *Scheme 3.2*).



*Scheme 3.2: Reaction of oxiranecarbonitrile (**94**) with phosphate ending in a cross-aldol reaction of glyceraldehyde-2-phosphate (**45**) and glycolaldehyde-phosphate (**16**), yielding pentose-2,4-bisphosphates, as described by Eschenmoser*[301].

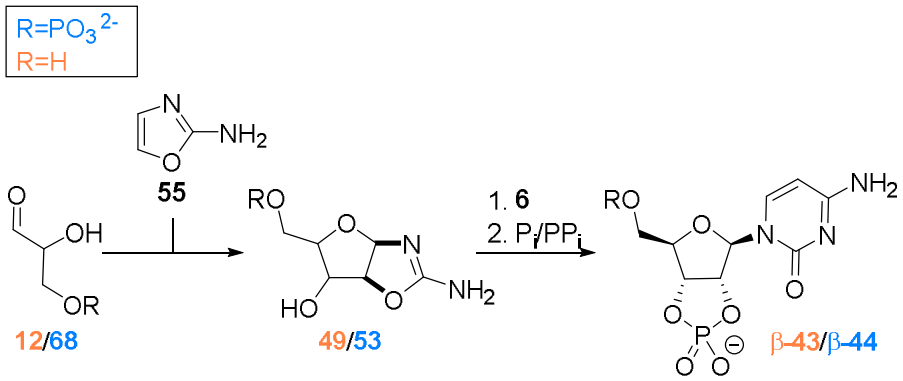
Milder conditions (1mM **16**, 4mM **45**, pH 7) can also be used to obtain a good yield of pentose-2,4-diphosphates (25%) again with a preference for the ribose variant **17** (48% of the pentose-2,4-diphosphates), when in the presence of mineral layers (mangalite, 30mM, 6 days[302]). The preference for the *ribo*- variant is attributed to the minimised 1,5 interactions in the transition state[273]. In spite of the selective synthesis of a ribose sugar, starting with these reactants prevents furanosyl formation, due to a phosphate group masking the C4′-hydroxyl, and leads to phosphorylation in the non-natural positions (2′ and 4′ instead of the desired 3′ or 5′). Another problem in this synthesis concerns the prebiotic plausibility of the epoxide **94**, which is formed from pure acrylonitrile (**81**) with NaOCl (1.8M)[301]. Pitsch *et al.* point out this problem by stating that "the question as to whether oxiranecarbonitrile [**94**] itself has the potential of having been a prebiological natural constituent remains unanswered"[301]. Reacting **81** (2.6-3.3M) with a prebiotically plausible oxidant (*vide supra*, *Section 2*) such as H_2O_2[53,293,295,296] (1 eq; pH 7.3-7.5, 35°C. in H_2O), yields glycidamide (**95**) in good yield (65-70%) with no trace of epoxide **94** (*Scheme 3.3*).



*Scheme 3.3: Reaction of acrylonitrile (**81**) with NaOCl forming oxiranecarbonitrile (**94**), or with H_2O_2 forming glycidamide (**95**). **95** is itself a hydrolysis product of the nitrile moiety of **94**.*

Unlike epoxide **94**, glycidaldehyde (**69**) can be accessed in a prebiotically plausible manner, from the oxidation of acrolein (**1**) with H_2O_2 (as described in detail in *Section 2*). As discussed previously, the constitutional relationship between **69** and glyceraldehyde (**12**) makes **69** an ideal candidate for the study of its involvement in nucleotide synthesis (*vide supra*, *Section 2*). Aldehyde **12** is a

known nucleotide precursor[33,272] used in previous nucleotide synthesis *via* masked-aldol reaction with 2-aminooxazole (**55**) and cyclisation to form key intermediates pentose aminooxazoline[33] (**49**; *Scheme 3.4*). The terminal phosphorylated variant of **12**, glyceraldehyde-3-phosphate (**68**) has also been proposed as a nucleotide precursor by similar reaction with **55** to yield 5′-phosphorylated pentose aminooxazolines (**53**), as described by Anastasi *et al.*[283] (*Scheme 3.4*). In their synthesis, Anastasi *et al.* remark on the need for a prebiotic pathway to **68**, which they use as a way to incorporate phosphate into nucleotides, but **68** is also a key molecule in its own right in other areas of biology, as it is an essential metabolite in many metabolic pathways such as glycolysis[22].

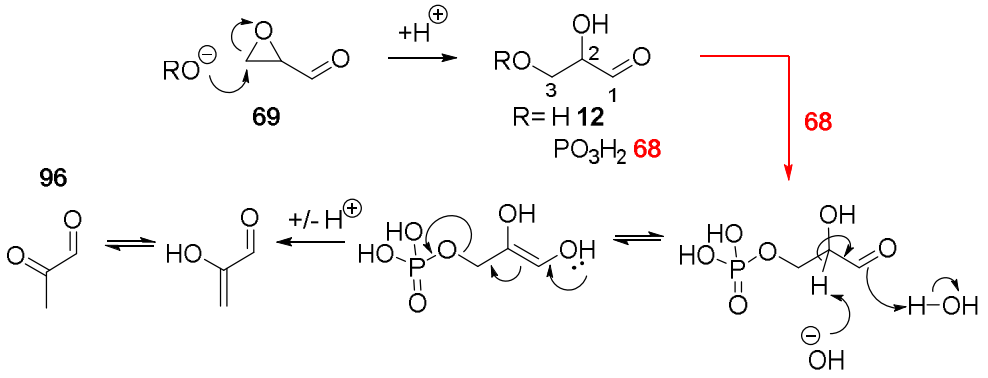


*Scheme 3.4: Prebiotic synthesis of activated pyrimidine ribonucleotides as described by Powner et al.[33] (orange) and Anastasi et al.[283] (blue) showing the synthesis of pentose aminooxazolines (**49**) or 5′-phosphorylated aminooxazolines (**53**) from 2-aminooxazole (**55**) and glyceraldehyde (**12**) or glyceraldehyde-3-phosphate (**68**), respectively. Reactions of aminooxazolines with cyanoacetylene (**6**) followed by dry-state phosphorylation then yields cytidine nucleotides β-**43**/β-**44**.*

The good yield of glycolaldehyde-phosphate (**16**; 47%, *vide supra*) achieved by Eschenmoser and co-workers[301] suggests efficient epoxide opening reactions are feasible in aqueous condition. Given the appreciable stability of glycidaldehyde (**69**) in water at neutral pH (*vide supra, Section 2*), and the higher nucleophilicity of phosphate dianions (the pK_as of phosphate - pK_{a1} 2.1, pK_{a2} 7.2 and pK_{a3} 12.6) compared with water (pK_a 14), we hypothesised that by harnessing the nucleophilic aqueous phosphorylation demonstrated by Eschenmoser, the

selective aqueous phosphorylation of **69** would furnish key nucleotide precursor glyceraldehyde-3-phosphate (**68**). It has been suggested that the increased electrophilicity at the α-position of α,β-epoxide carbonyls should lead to preferential nucleophilic attack at the α-position[303]. However, the hydration of the aldehyde moiety of **69** in water, leading **69** to sit as the hydrate **69**·H_2O (<95%; *Section 2*) led us to expect loss of this preference for the α-position and instead selective phosphorylation at the less hindered C3 position. This nucleophilic phosphorylation of **69** at the terminal C3 would limit the need for high concentrations, elevated temperatures or pre-activation of phosphate while bringing us closer to producing the natural isomers of nucleotides (3′ and 5′ phosphates) prebiotically.

An initial study of the reaction of glycidaldehyde (**69**) in phosphate buffer (equimolar ratio of **69** to P_i at 250mM concentration each, pH 7) revealed the formation of a mixture of compounds, including hydrolysis to glyceraldehyde (**12**) in 9% yield after 7 days (based on ^{1}H NMR integration, percentages reported as relative to identifiable species in solution, *Scheme 3.5*). Alongside hydrolysis, ^{31}P NMR analysis revealed a triplet (6.41 ppm, J = 6.6 Hz), that was interpreted as the C3-phosphate of glyceraldehyde-3-phosphate (**68**), which was then confirmed by spiking with an authentic sample (Appendix – *Figure 11.8*). The observed yield of **68** was 15% yield based on ^{1}H NMR integration.



*Scheme 3.5: Proposed mechanism for the opening of glycidaldehyde (**69**) with hydroxide, leading to glyceraldehyde (**12**) or with phosphate, leading to glyceraldehyde-3-phosphate (**68**). Formation of **68** is followed by elimination of phosphate via E1cB resulting in methylglyoxal (**96**) production.*

While initial results showed phosphate incorporation at the desired C3 position, the yield of **68** was low. The rate of nucleophilic attack by phosphate at the terminal C3 atom of epoxide **69** was believed to be proportional to the concentration of the nucleophilic phosphate dianion. Therefore, to the test the effect of pH (and dianion concentration) on the phosphorylation of **69**, we next incubated **69** in P_i buffer at a range of pH (5 to 8), crossing the second pK_a of P_i (pK_{a2} 7.2), as well as in the presence of various amounts of P_i (1:1 to 4:1 P_i:**69**; *Figure 3.3*). The higher pH reactions (pH 8) with an excess of P_i (4:1 or 2:1, P_i:**69**) showed the highest yields of **68**, of about 20% (assessed by ^{1}H NMR integration relative to other compounds present in the sample, *Figure 3.4*). However, the increased alkaline environment also increased our product **68**'s propensity to degrade into methylglyoxal (**96**) by E1cB elimination, as observed in the study by Gauss *et al.*[304], leading to lower yields than desired. Gauss *et al.* report a near quantitative elimination of **68** to **96** after 4 days at neutral pH (pH 6.8, 2.3mM in water, RT), which corroborates the elimination observed in our experiments.

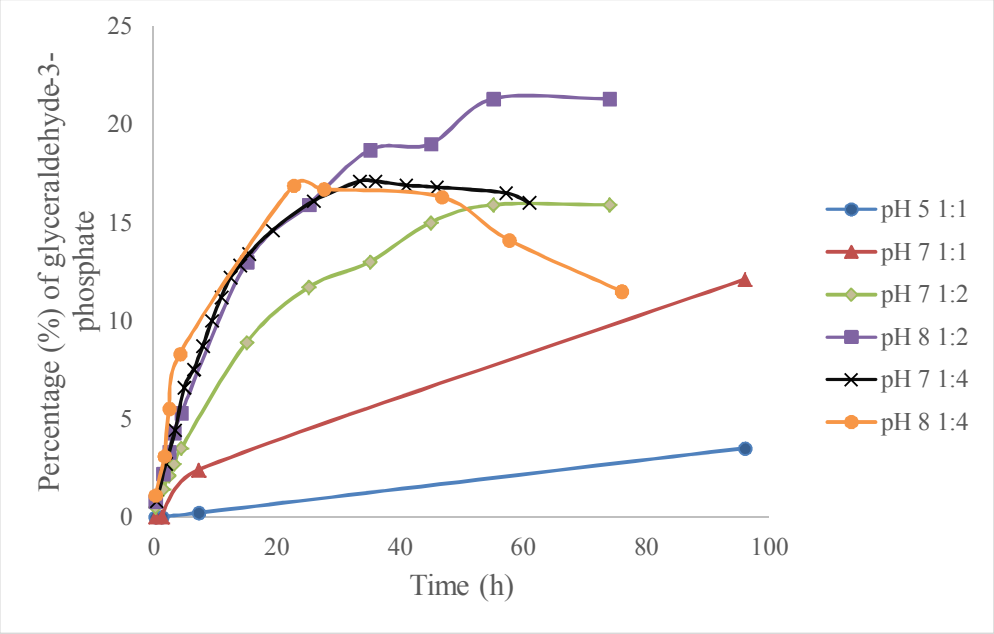


*Figure 3.3: Comparison of glyceraldehyde-3-phosphate (**68**) production at varying pHs and P_i concentrations over time. The reactions of glycidaldehyde (**69**; 250mM) in water at RT were analysed periodically by ^{1}H NMR spectroscopy. Percentages are reported as relative to identifiable species in solution. Ratios given represent **69**:P_i.*

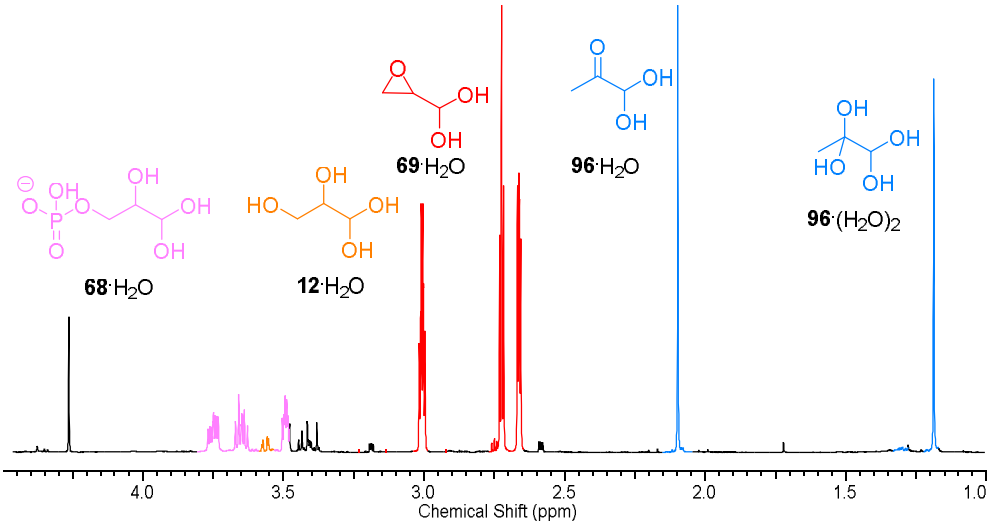


*Figure 3.4: ^{1}H NMR spectrum (600 MHz, H_2O/D_2O 9:1) of glycidaldehyde (**69**) (0.25M) in P_i buffer (4 eq; pH 7) after incubation at RT for 36 h. Two methyl resonances corresponding to methylglyoxal (**96**) mono- and dihydrate are observed (blue). The other aldehydes in solution: **69**, glyceraldehyde (**12**) and glyceraldehyde-3-phosphate (**68**) sit as their hydrate form (·H_2O).*

The observed pH-dependent behaviour of glyceraldehyde-3-phosphate (**68**) shows a significant increase in stability at acidic pH (pH<4). However, these conditions hinder nucleophilic phosphorylation due to a lower concentration of phosphate dianion, thereby leading to loss of efficiency, as observed in our pH 5 experiments (*Figure 3.3*, *vide supra*). The limited phosphorylation allowed more opportunity for both hydrolysis of the starting material **69** to **12** as well as elimination of the desired product **68** to **96**. The lack of phosphorylation potential at low pH and the instability of **68** at neutral to basic pH effectively render **68** unavailable in a prebiotic context.

When taking into account the phosphorylation pathway as a whole, from the formation of **68** to its subsequent phosphate elimination to **96**, we found that the proportion of products derived from terminal nucleophilic phosphorylation is significant (combined yield of **68** and **96** is around 60%). The highest yields are achieved in the higher pH reactions with an excess of phosphate (pH 8, 4:1 or 2:1 P_i:**69**, 61% after 74 h and 53% after 76 h, respective combined yields, *Figure 3.5*).

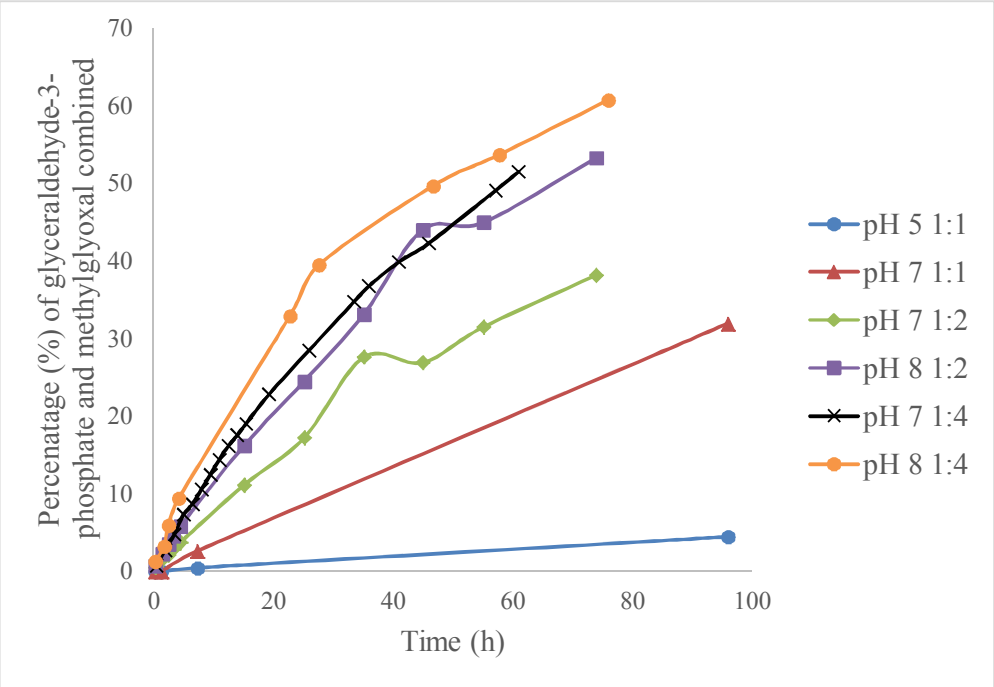


*Figure 3.5: Comparison of total glyceraldehyde-3-phosphate (**68**) production (including its degradation product methylglyoxal (**96**)) at varying pHs and P_i concentrations over time. The reactions of glycidaldehyde (**69**; 250mM) in water at RT were analysed periodically by ^1H NMR spectroscopy. Percentages are reported as relative to identifiable species in solution. Ratios given represent **69**:P_i.*

The instability of glyceraldehyde-3-phosphate **68** and consequently its unavailability for any future prebiotic reactions make it a poor candidate for a phosphorylated feedstock molecule as envisaged by Anastasi *et al*[283]. However, the efficiency of the reaction with regards to initial phosphate trapping in water makes glycidaldehyde (**69**) a good potential candidate for other phosphorylation reactions. This excellent selectivity for the formation of a primary phosphate, as seen in 5'-phosphorylated nucleotides, leads us to consider alternative routes using **69** to synthesise ribonucleotides through 5'-phosphorylation in water.

4. Prebiotic synthesis of aminooxazolines through glycidaldehyde

4.1. Building aminooxazolines

There have been several major stumbling blocks in the synthesis of ribonucleotides such as the synthesis of the sugar, the bases and the subsequent glycosidation reaction of the two, which all come with their share of major problems, namely low yields, lack of selectivity and thermodynamic obstacles[166] (*vide supra*, *Section 1.2.2-1.2.4*). A solution to the glycosidation reaction was potentially found, however, when Sanchez and Orgel discovered ribocytidine (**10**) could be formed from the addition of cyanamide (**4**) to ribose (**9**), followed by the addition of cyanoacetylene (**6**)[270] (albeit the α-anomer α-**10**, a point to be returned to later). The studies revealed the initial formation of riboaminooxazoline (*ribo*-**49**) when **9** is reacted with an excess of **4** in aqueous ammonia (3.5%), leading to an exothermic reaction of temperatures up to 85°C. Upon cooling, *ribo*-**49** crystals can be collected (up to 94% yield when using methanol as an anti-solvent). Another interesting finding was the greater efficiency and speed of **9** reacting with **4** compared to other pentose sugars, which was attributed to **9**'s favoured predisposition to sit as its aldehydic form compared to other pentoses[209] (*Figure 4.1*). This increase in reaction speed effectively means that ribose (**9**) is less affected by the depletion of cyanamide (**4**) from solution, due to **4** dimerisation and hydrolysis to urea (**28**; the half-life of 1M **4** is in the order of half a day at pH 8, 1M P_i buffer, 65°C)[223]. Furthermore, the favoured reaction of **9** with **4** compared to other pentoses provides a means to gain chemical control in sugar mixtures, which is further enhanced by the selective crystallisation of *ribo*-**49** from solution (over other pentose **49**)[305]. Finally, formation of *ribo*-**49** increases the stability of ribose 70-fold, making *ribo*-**49** more prebiotically available than the free sugar form **9**[305].

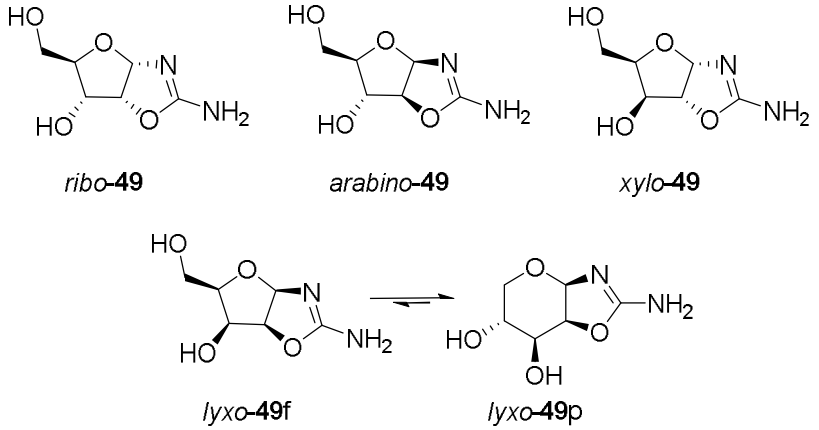
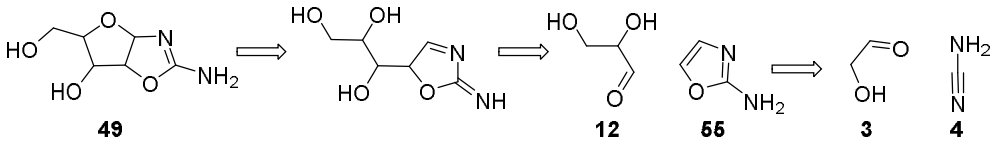


*Figure 4.1: The five stereoisomers of pentose aminooxazolines (**49**). Ribo-**49**, arabino-**49** and xylo-**49** preferentially sit in the furanosyl form in water, however lyxoaminooxazoline (lyxo-**49**) favours the pyranosyl form 5:1, lyxo-**49**p:lyxo-**49**f*[272,306].

The synthesis of aminooxazolines **49** is an inventive way to by-pass the traditional glycosidation reaction[208] while selectively forming furanosyl-nucleoside precursors[306] (with the exception of *lyxo*-**49** that preferentially forms the pyranosyl form due to "steric crowding on the β-face" in the furanosyl form[306]; a point to be returned to later). However, **49** synthesis still suffers from stumbling blocks such as obligate C1′-C2′ *cis*-stereochemical relationship in **49**, which, once reacted with cyanoacetylene (**6**) to complete the pyrimidine nucleobase moiety, leads to products one stereochemical inversion away from the desired β-ribocytidine (β-**10**) (*ribo*-**49**, leading to α-**10**; or *arabino*-**49**, leading to β-arabinocytidine (β-**51**), but not the canonical β-**10**). Another problem is the use of pure pentose sugar as starting material for **49** synthesis (*vide supra*, *Section 1.2.2*). Indeed, one of the most promising prebiotic ribose (**9**) syntheses is from the $Zn(Pro)_2$ catalysed cross-aldolisation reaction of glycolaldehyde (**3**) and glyceraldehyde (**12**) to form mainly pentose sugars (62% yield) with a preference for **9** (20% overall yield)[186v]. However, this synthesis is disrupted by the presence of cyanamide (**4**), leading to competing chemistry (a point to be returned to later, *Section 5*), and thereby making **9** synthesis incompatible with **49** synthesis.

[v] Other prebiotic attempts at ribose (**9**) synthesis are described in *Section 1.2.2.*

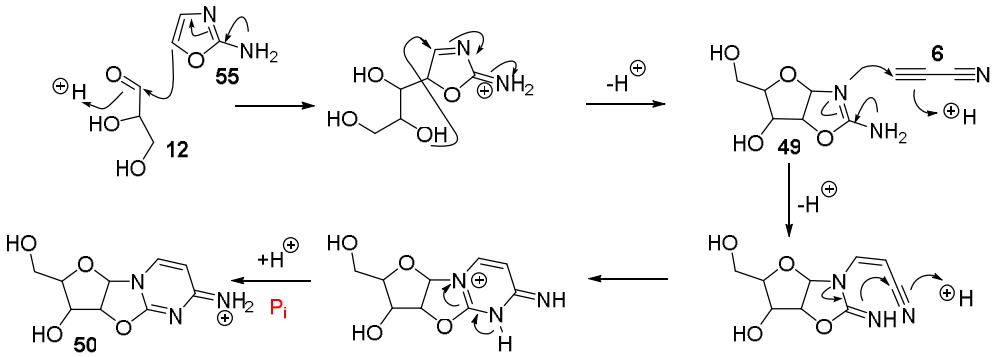
The Sutherland group envisaged a new retrosynthetic analysis to the nucleoside problem, by-passing the need for the problematic synthesis of the pentose sugar **9**[272] (*Scheme 4.1*). A central compound from this new route is the sugar and base synthon 2-aminooxazole (**55**), generated from the reaction of glycolaldehyde (**3**) and cyanamide (**4**) in P_i buffer (>80% yield)[33]. The subsequent one-pot reaction with glyceraldehyde (**12**) yields *rac*-pentose aminooxazolines (**49**) in 50% yield (over two steps) with a diastereotopic selectivity of *ribo*-**49**>*arabino*-**49**>>*lyxo*-**49**>*xylo*-**49** (*Figure 4.1*).



*Scheme 4.1: Retrosynthetic analysis of pentose aminooxazolines (**49**) generating C_1, C_2 and C_3 carbon units.*

The reaction developed by the Sutherland group shows a strong (78%) stereoselectivity for the [*3R*,*4R*]-**49** (*ribo*- and *arabino*-), both of which are one stereochemical inversion away from β-ribofuranosyl-conformation. This new synthesis is highly selective for the canonical furanosyl conformer (only the *lyxo*-variant sits preferentially in the pyranosyl form, 5:1 pyranosyl/furanosyl[272]), as well as for the *ribo*- configuration (44% of the pentose sugars compared to 30% in the glycolaldehyde (**3**) and glyceraldehyde (**12**) cross-aldol reaction[186] or 22% in the aldomerisation of glycolaldehyde phosphate (**16**)[273], *vide supra*, *Section 1.2.2*).

The *arabino*-**49** and *ribo*-**49** rich reaction can be further enriched in *arabino*-**49** by selective crystallisation of *ribo*-**49** from solution[272]. The *arabino*-**49** in solution is then reacted with cyanoacetylene (**6**) similarly to Sanchez and Orgel[270], yielding arabinose furanosyl-2,2′-anhydrocytidine (*arabino*-**50**; *Scheme 4.2*) in a phosphate mediated reaction (a point to be returned to later, *Section 6*).



*Scheme 4.2: Mechanism of aminooxazolines (**49**) formation from the masked-aldol reaction of glyceraldehyde (**12**) and 2-aminooxazole (**55**) and the subsequent reaction of cyanoacetylene (**6**) with **49** to yield furanosyl-2,2′-anhydrocytidine (**50**). The last step is mediated by P_i (red) which acts as a pH buffer.*

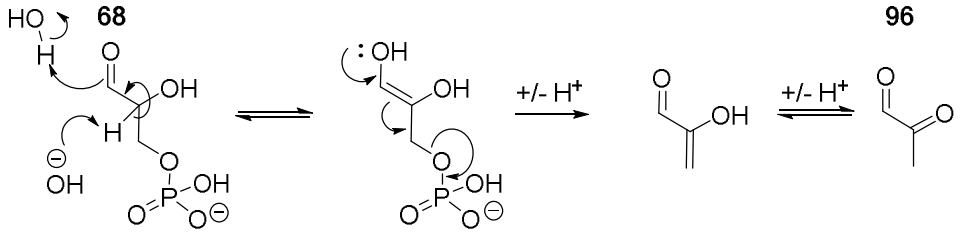
Upon evaporation and heating with P_i and urea (**28**), *arabino*-**50** is then selectively phosphorylated at the C3′-hydroxyl position (*vide supra*, *Section 3*, *Figure 3.2*), followed by cyclisation onto the C2′ carbon atom. This leads to stereochemical inversion at the 2′-stereocentre, yielding the canonical nucleotide β-ribofuranosylcytidine-2′-3′-cyclic phosphate (β-**43**; 46% yield)[33]. Finally, they observed that irradiation of a solution of β-**43** resulted in partial hydrolysis to β-ribouridine-2',3'-cyclic phosphate (β-**57**), whilst destroying all other pyrimidine side products resulting from previous steps, thereby enriching the solution in the two canonical activated pyrimidine ribonucleotides and minimising unwanted by-products.

The chemistry described above solves the problems associated with Sanchez and Orgel's route to aminooxazolines (**49**), namely the stereochemistry issue (product are one stereochemical inversion away from the canonical β-**10**: α-**10** or β-**51** when starting with *ribo*-**49**, or *arabino*-**49** respectively) and avoids the need for ribose (**9**); thereby putting **49** back at the centre of prebiotic synthesis of nucleotides. This new method, involving common C_1, C_2 and C_3 building blocks opens up possibilities of using variants on these building blocks to yield modified nucleotides. One such variation of the reaction is that of glyceraldehyde-3-phosphate (**68**) and 2-aminooxazole (**55**) as described by Anastasi *et al.*[283], yielding 5′-phosphorylated aminooxazolines (**53**; *Table 4.2*, entry 4; *vide infra*).

We envisaged this new flexible chemistry could be utilised with our epoxide, glycidaldehyde (**69**), in order to potentially incorporate phosphate into our building blocks along the way, without the need for heating and dry-state phosphorylation.

4.2. Use of glycidaldehyde in aminooxazoline synthesis

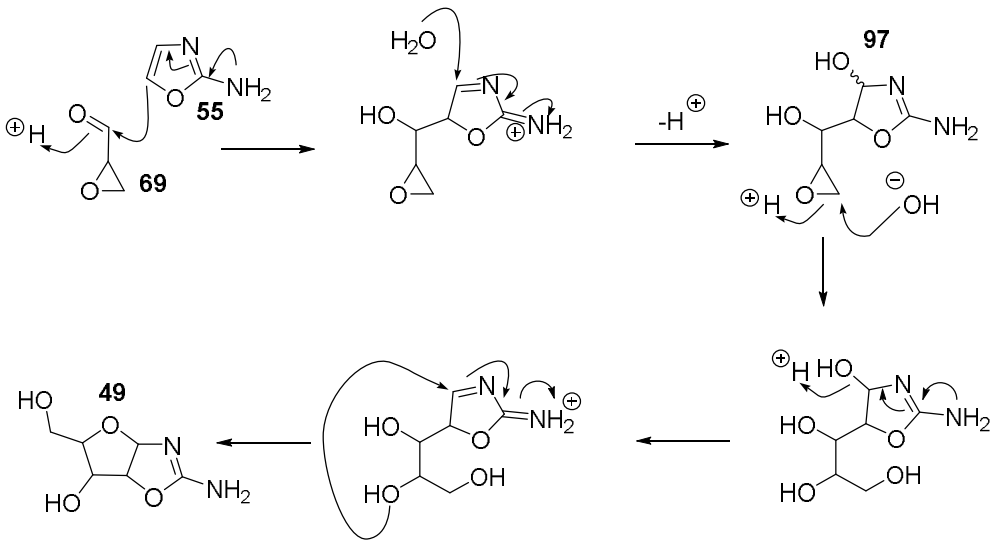
Previously we discussed the reaction of glycidaldehyde (**69**) with phosphate to yield glyceraldehyde-3-phosphate (**68**), as a way to tether phosphate selectively and efficiently in water (*vide supra*, *Section 3*). However, the instability of **68** makes it difficult to establish **68** as a key compound in nucleotide synthesis. This instability is due to phosphate elimination in **68** yielding methylglyoxal (**96**) through E1cB elimination and is initiated by enol formation from the aldehyde moiety as depicted in *Scheme 4.3*. The detrimental E1cB elimination would be prohibited if the aldehyde moiety of **69** was not available. Accordingly, we hypothesised that by reversing the sequence of reactions in the synthesis of 5′-phosphorylated aminooxazolines (**53**), aldol then phosphorylation, we might retain the epoxide's potential for terminal phosphorylation while by-passing E1cB elimination.



*Scheme 4.3: Mechanism of glyceraldehyde-3-phosphate (**68**) E1cB elimination leading to the formation of methylglyoxal (**96**).*

We proposed that due to the constitutional similarity to glyceraldehyde (**12**), **69** might undergo masked-aldol reaction with 2-aminooxazole (**55**; *Scheme 4.4*). We

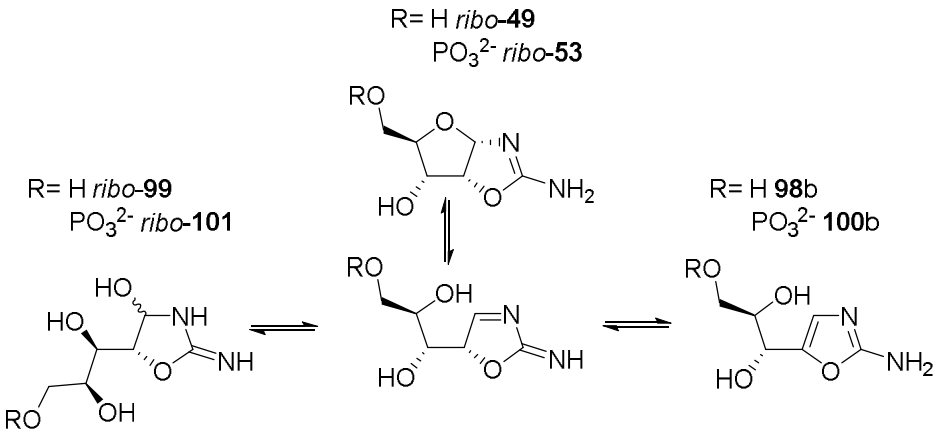
hypothesised that a masked aldol-type reaction may occur initially to form oxazoline **97**. Hydration of **97** would then provide a mechanism for the formation of known pyrimidine nucleotide precursor pentose aminooxazoline (**49**), by-passing the requirement for **12** (a point to be returned to, *Section 5*) during nucleotide assembly (*Scheme 4.4*). We suspected that 5-*exo-tet* cyclisation and epoxide opening of **97** would be prohibited due to poor stereoelectronic overlap of $n_{O1'} \rightarrow \sigma^*_{C2\text{-}O}$. Interestingly, the inhibition of intramolecular rearrangement of **97** opens the exciting possibility to use the epoxide's excellent selectivity and efficacy of nucleophilic insertion (*vide supra*, *Section 3*) to regiospecifically activate nucleotide precursor **49** with 5′-modification, including 5′-phosphorylation.



*Scheme 4.4: Proposed mechanism of pentose aminooxazolines (**49**) formation from the masked-aldol reaction of glycidaldehyde (**69**) and 2-aminooxazole (**55**), via oxazoline **97**.*

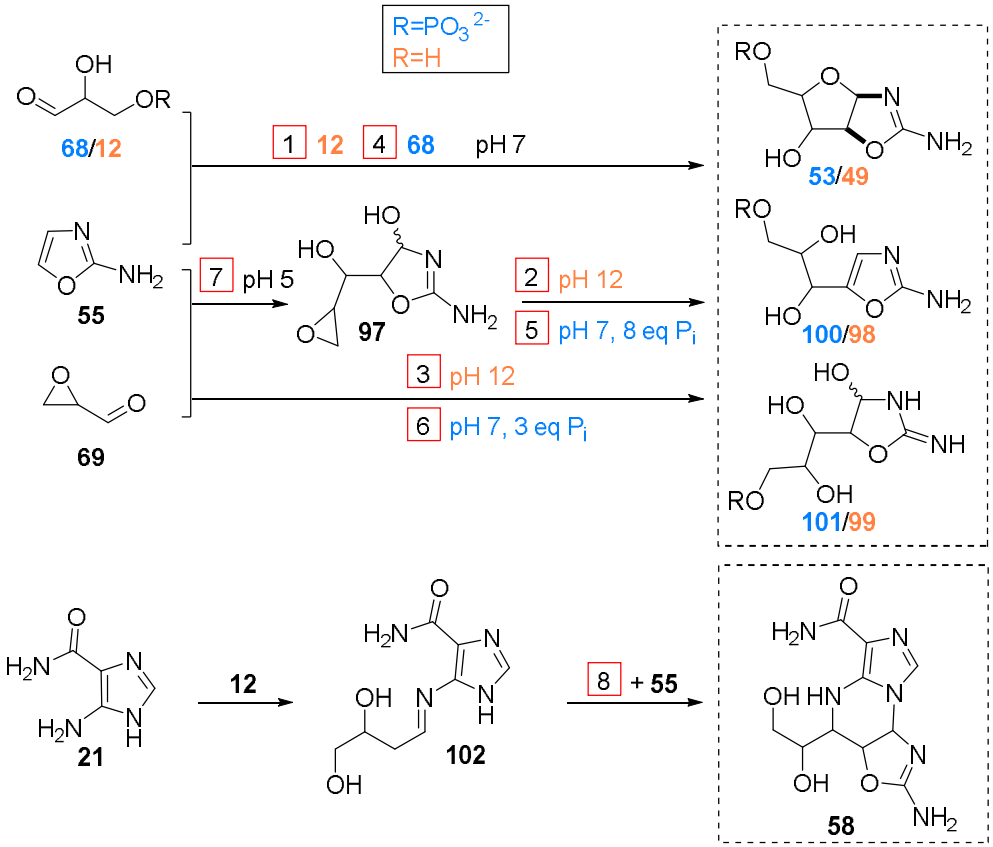
Before attempting variations of the reaction developed by the Sutherland group (*vide supra*), we decided to repeat the reaction of synthetic 2-aminooxazole (**55**) with glyceraldehyde (**12**) in order to comparably analyse the diastereomer selectivity with any future reactions we might undertake (*vide infra*, *Section 4.2.* and *Section 4.3.*). Our results showed that a reaction of **12** (140mM) and **55** (168mM) in water yielded pentose aminooxazolines in 74% yield (^{1}H NMR

analysis with an internal standard, *Table 4.1*, entry 1) with a 44:34:13:9 *ribo*/*arabino*/*lyxo*/*xylo*-**49** selectivity (*Table 4.2*, entry 1), in accordance with the reported ratios[272]. Some of the remaining material is notably the open chain variants of aminooxazolines: the aromatic form **98** and the hydrate form **99** (*Scheme 4.5*)[vi].



*Scheme 4.5: Open-chain versions of pentose aminooxazolines (**49**) and aminooxazolines 5'-phosphate (**53**): aromatic form **98/100** and hydrate form **99/101**, as demonstrated on the ribo- diastereomer.*

[vi] The open-chain versions of aminooxazolines can lead to anomerisation of the *cis*-C1/C2, as will be discussed in *Section 6*.



*Scheme 4.6: Reactions analysed in Tables 4.1, 4.2 and 4.4. Table entry numbers shown next to the corresponding reaction (red box). Outcome of the reactions shown in dotted box. 1. Reaction of glyceraldehyde (**12**; 140mM) and 2-aminooxazole (**55**; 1 eq), pH 7, RT. 2. Hydrolysis of adduct **97** (prepared as described in 7.), at pH 12, RT. 3. Reaction of glycidaldehyde (**69**; 128mM) and **55** (1.1 eq), pH 12, RT. 4. Reaction of glyceraldehyde-3-phosphate (**68**; 60mM) and **55** (1 eq), pH 7, RT. 5. Phosphorylation of **97** (prepared as described in 7.), at pH 7 (8 eq P_i). 6. Reaction of **69** (128mM), **55** (1.1 eq) and P_i (3 eq), pH 7, RT. 7. Reaction of **69** (128mM) and **55** (1.3 eq), pH 5, RT, to form adduct **97**. 8. Reaction analysed based on ratios adapted from Powner et al.[274] (in order to add up to 100).*

Entry n°	Reaction	^{1}H NMR integration relative to internal standard (%)			
		49/**53**[a]	**99**/**101**[b]	**98**/**100**[c]	Total
1	**12** + **55**	74	3	0	77
2	**97** + OH^-	70	0	0	70
3	**69** + **55** to **49**	15	0	0	15
4	**68** + **55**	25	11	3	39
5	**97** + P_i	30	3	12	45
6	**69** + **55** + P_i	21	0	17	38

Table 4.1: Absolute yields of various aminooxazoline assembly reactions. Yields were calculated by ^{1}H NMR spectra integration (600 MHz, H_2O/D_2O, 9:1) relative to an internal standard (KHP). Reactions described in Scheme 4.6.

[a]Entries 1-3: integration of the H-(C1′) doublets of aminooxazolines (**49**); entries 4-6: integration of the H-(C1′) doublets of aminooxazolines 5′-phosphate (**53**; can contain minor **49**). [b]Entries 1-3: integration of the H-(C1′) signals of open chain hydrates **99**; entries 4-6: integration of the H-(C1′) signals of 5′-phosphate open chain hydrates **101**. These numbers might be inflated due to overlap of the 5.5 ppm region with the hydrate of 2-aminooxazole (**55**): **55**·H_2O. [c]Entries 1-3: integration of the H-(C1′) signals of open chain aromatic species **98**; entries 4-6: integration of the H-(C1′) signals of 5′-phosphate open chain aromatic species **100**. These numbers might be inflated due to overlap of the 6.5 ppm region with **55**.

Entry n°	Reaction	^{1}H NMR integration relative to other aminooxazolines integrated (%)				
		ribo	*arabino*	*lyxo* (f)	*lyxo* (p)	*xylo*
1	**12** + **55**	44	34	3	10	9
2	**97** + OH^-	42	16	3	13	26
3	**69** + **55** to **49**	39	16	4	15	26
4[a]	**68** + **55**	63	23	NA	NA	14
5*	**97** + P_i	45	16	NA	NA	39
6[b]*	**69** + **55** + P_i	45	11	NA	NA	44
7[c]	**69** + **55** to **97**	40	13	23	0	24
8	**102** + **55**	13	6	72	NA	9

*Table 4.2: Stereoselectivity in various aminooxazoline assembly reactions. Unless otherwise stated, ratio calculation for the four stereoisomers of aminooxazolines (or in the precursor oxazoline **97**) was performed by ^{1}H NMR spectra integration (600 MHz, H_2O/D_2O, 9:1) as a percentage of the relevant H-C1′ peaks and are reported as a mean over 3 reactions. Reactions described in Scheme 4.6. Entries 1-3: Ratios of H-(C1') doublets of aminooxazolines (**49**). Entries 4-6: Ratios of the H-(C1') doublets of aminooxazolines 5'-phosphate (**53**). Entry 7: Ratios of the H-(C1') doublets of adduct **97**. Entry 8: Ratios of the aminooxazolines **58**.*

[a]Ratios were calculated from one reaction. [b]Ratios calculated from two separate reactions, averaging the H-C1′ and H-C2′ integrations (due to poor resolution of baseline leading to higher error). [c]The four distinct doublets interpreted as H-C1′ of **97** leading to the four diastereomers of aminooxazolines were assigned as leading to *ribo*-, *lyxo*-, *xylo*- and *arabino*-isomers, from left to right (*Figure 4.2*, inset; *vide infra*). *Ratios only include discernible cyclised aminooxazolines and were calculated after purification by OH^- Dowex® column.

We then went on to study the reaction of 2-aminooxazole (**55**) with glycidaldehyde (**69**). Initially, two parallel equimolar reactions of **55** and **69** were monitored at RT in water, at pH 5 and 7 respectively. The ^{1}H NMR analysis of the pH 5 reaction showed a complex mixture of products including four discernible doublets in the 5.70-5.65 ppm region (5.69, d, J = 2.5 Hz; 5.68, d, J = 2.7 Hz; 5.64, d, J = 2.6 Hz; 5.62, d, J = 2.4 Hz), which we interpreted as being the H-C1′ of oxazoline **97** (*Figure 4.2*, ***97***[vii]). This interpretation was corroborated by 2D-NMR analysis, providing evidence that these four protons were each coupled to an extended (>C_2) resonance system.

[vii] NMR spectra of the reaction over time can be found in the Appendix – *Figure 11.1*.

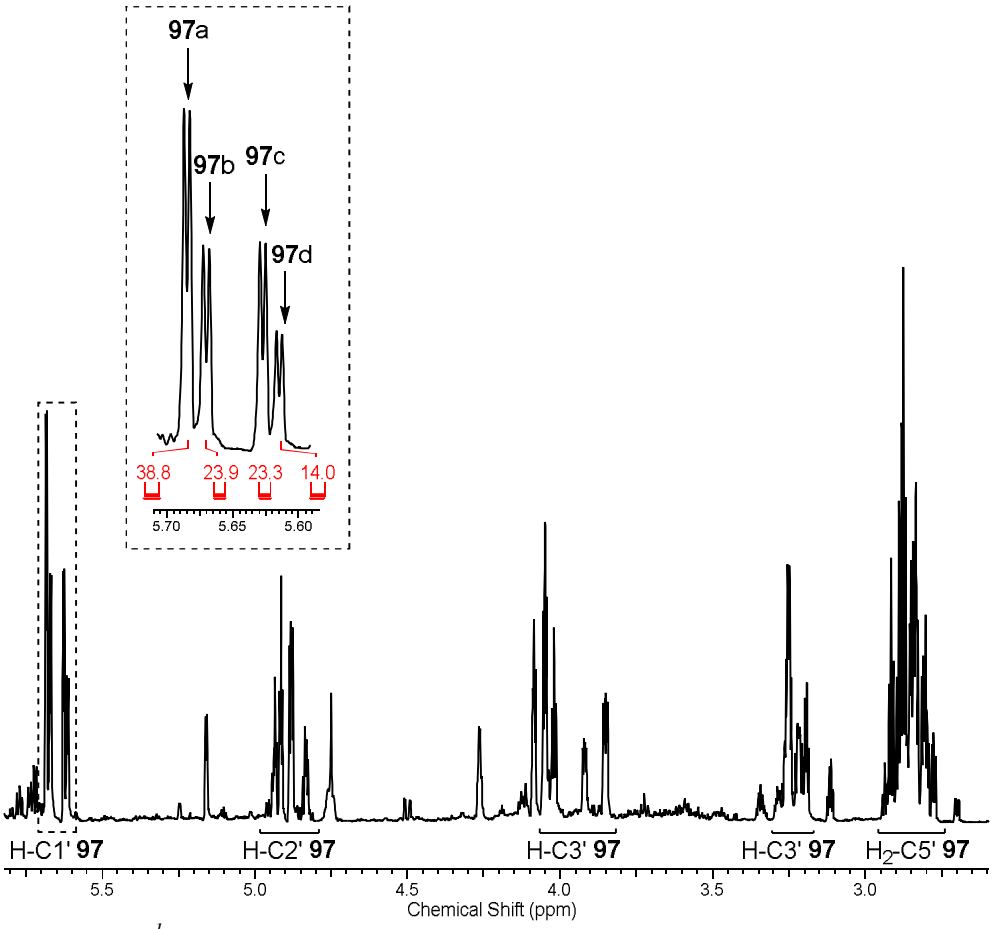


*Figure 4.2: ^{1}H NMR spectrum (600 MHz, H_2O/D_2O 9:1, 5.76-2.60 ppm) of the reaction of glycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.1 eq) at pH 5 after 15 h at RT, showing the four discernible doublets in the 5.7 ppm region, attributed to oxazolines **97**. Inset (5.71-5.59 ppm): Integration of the four H-C1' of **97**.*

The ^{1}H NMR analysis of the pH 7 reaction showed the increase of four doublets in the 6.00-5.80 ppm region over time (6.00 ppm, d, J = 5.0 Hz; 5.96 ppm, d, J = 5.3 Hz; 5.85 ppm, d, J = 5.2 Hz; 5.79 ppm, d, J = 5.5 Hz), which were interpreted to be the H-C1′ protons of the pentose aminooxazolines (**49**)[viii]. Sample spiking with authentic standards of **49** confirmed that these new doublets were indeed the pentose aminooxazolines: riboaminooxazoline (*ribo*-**49**), arabinoaminoxazoline (*arabino*-**49**), xyloaminooxazoline (*xylo*-**49**) and the furanosyl form of lyxoaminooxazoline (*lyxo*-**49**f), (*Figure 4.3*); further reinforcing our original

[viii] NMR spectra of the reaction over time can be found in the Appendix – *Figure 11.2.*

interpretation that the 5.7 ppm doublets are the proposed intermediate oxazoline **97**. These reactions show evidence that glycidaldehyde (**69**) is undergoing masked-aldol reaction with 2-aminooxazole (**55**) *via* the aldehyde moiety, thereby showing promise concerning the resolution of our E1cB elimination problem mentioned above (*vide supra*, *Scheme 4.3*).

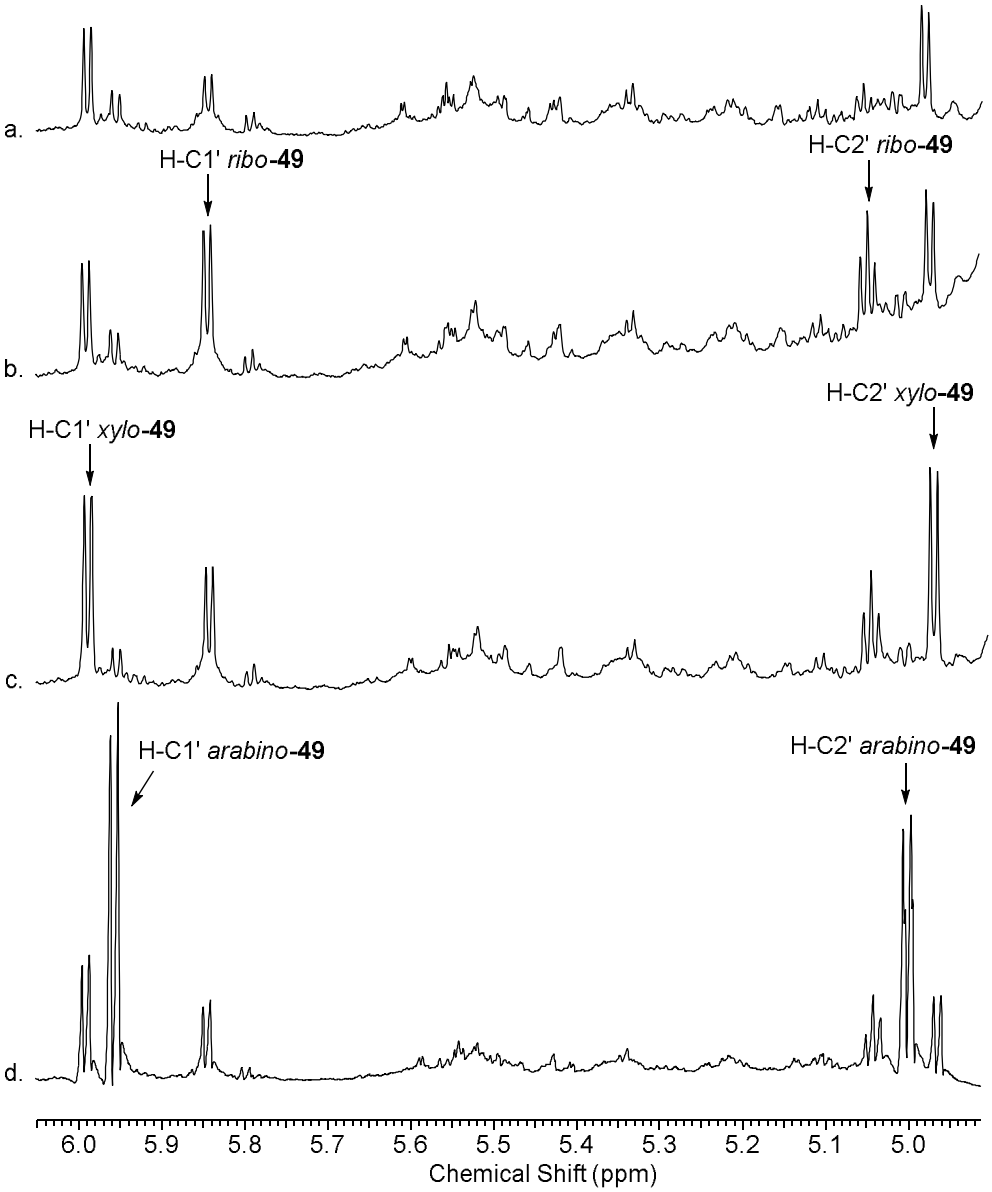


*Figure 4.3: ^{1}H NMR spectra (600 MHz, D_2O, 6.05-4.90 ppm) showing the H-C1' and H-C2' peaks for the four stereoisomers of the aminooxazolines (**49**) and subsequent signal amplification observed by spiking with synthetically prepared authentic standards of **49**. a. Reaction at pH 7 lyophilised after 28 h and dissolved in D_2O. b. Spiked with ribo-**49**. c. Spiked with xylo-**49**. d. Spiked with arabino-**49**.*

In order to gain more insight into the mechanism and selectivity of this reaction, the reaction was repeated at pH 5 and monitored until the system was stable, over 15 h at RT, during which the solution was adjusted to remain at pH 5 (pH fluctuations ±0.8 were observed; the system was deemed to be stable after the appearance of the desired doublets in the 5.7 ppm region with no obvious evolution of the ^{1}H NMR). The solution was then raised to pH 7 and the oxazoline **97** doublets were observed to slowly disappeared while the pentose aminooxazoline (**49**) doublets appeared. After 25 h at pH 7, the system stabilised with 40% of the adduct **97** having hydrolysed to **49** (based on ^{1}H NMR analysis, relative to the combined integration of **97** and **49** in solution; *Figure 4.4*, *a.*). Further conversion to **49** was observed after the solution was raised to pH 9 (*Figure 4.4*, *b.*). It was of note that different hydrolysis rates were observed for the four intermediate oxazolines **97**. The *ribo*- leading **97**a[ix] intermediate was slower to convert to **49**, such that after 20 h at pH 9, 60% of **97** was hydrolysed to **49**, with the **97**a doublet left at 5.35 ppm (*Figure 4.4*, *b.*). To complete the conversion of **97** to **49**, a higher concentration of hydroxide was required. Further incubation at pH 12 for 12 h led to full conversion to pentose aminooxazolines (**49**), with a final ratio order of *ribo*-**49**>*xylo*-**49**>>*arabino*-**49**>*lyxo*-**49** (*Figure 4.5*).

[ix] The **97**a intermediate was interpreted as leading to *ribo*-**49** due to spiking of the fully hydrolysed reaction (*vide infra*).

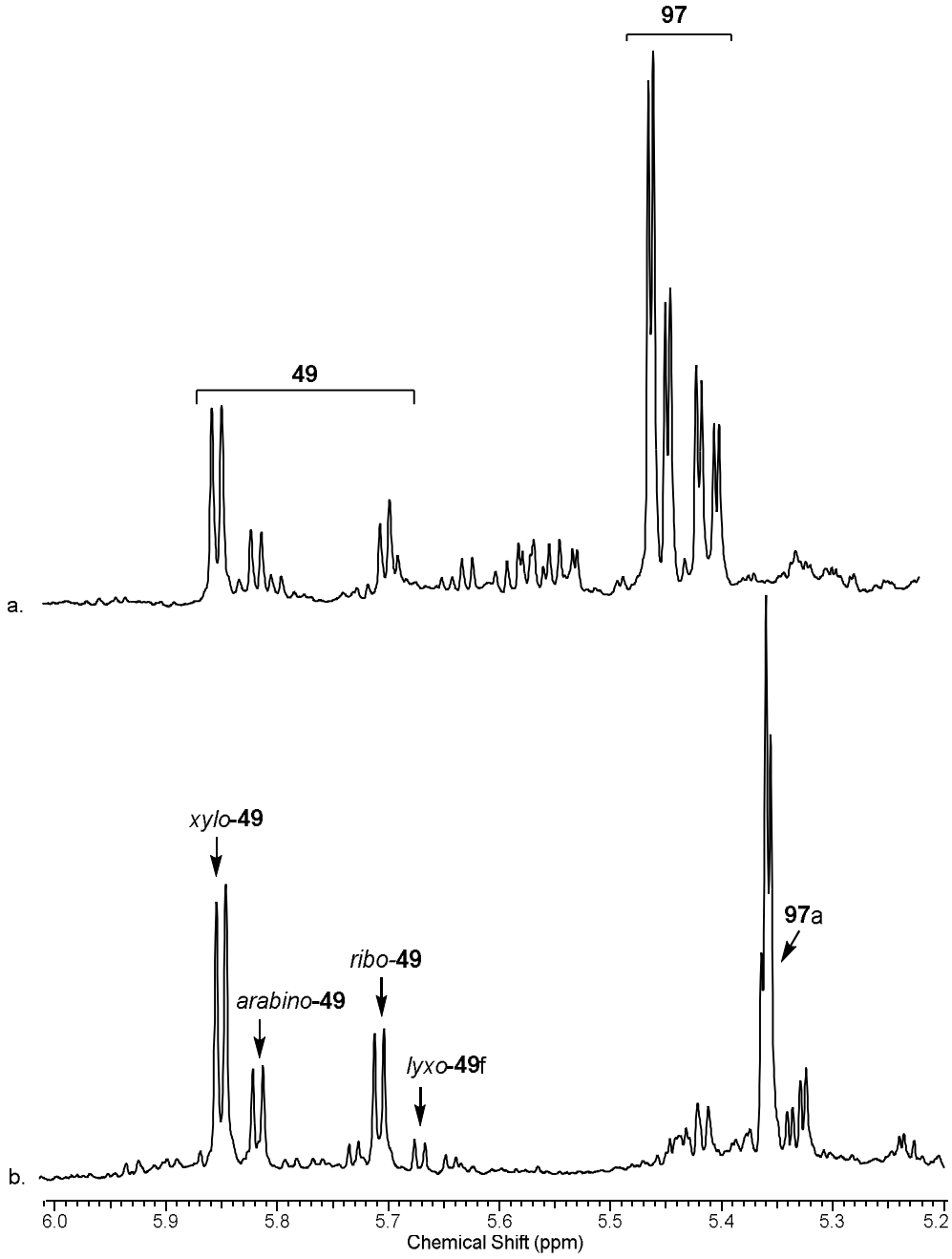


*Figure 4.4: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 6.00-5.20 ppm) showing the H-C1' for the four stereoisomers of the aminooxazolines (**49**) and the discernible proton peaks for adduct **97**. a. The reaction was carried out with equimolar concentrations of glycidaldehyde (**69**) and 2-aminooxazole (**55**) in water at pH 5 and at RT. The reaction was changed to pH 7 after 15 h. Spectrum acquired after 25 h at pH 7. b. The reaction was changed to pH 9 and adjusted thereafter. Spectrum acquired 20 h after pH change, showing conversion of most of the peaks for **97** to **49** peaks.*

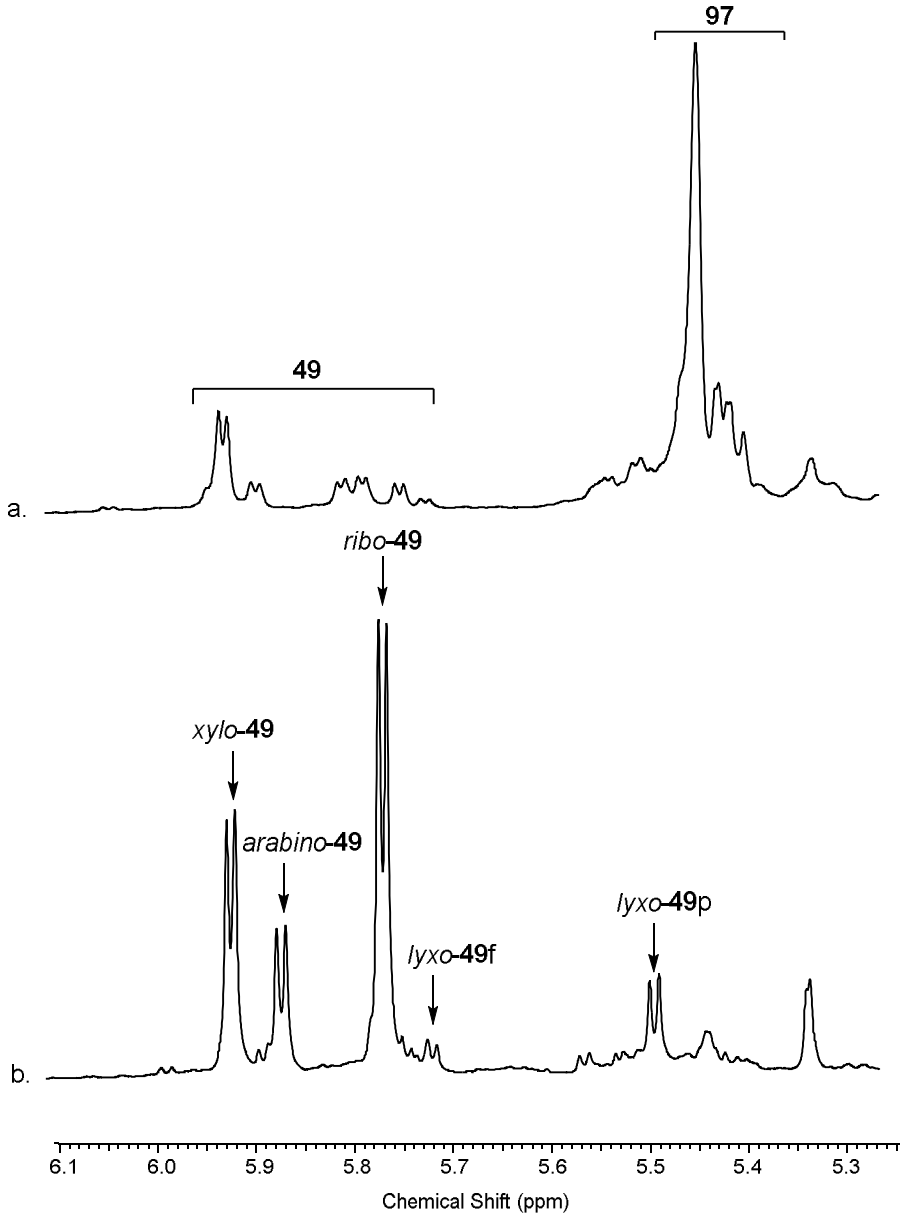


*Figure 4.5: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 6.10-5.25 ppm) showing the H-C1' for the four stereoisomers of the aminooxazolines (**49**) and the discernible proton peaks for adduct **97**. The reaction between glycidaldehyde (**69**) and 2-aminooxazole (**55**) was carried out and driven to completion, at pH 5 and at RT. a. The reaction was changed to pH 9, driving the reaction towards **49** formation. Spectrum acquired 24 h after pH change showing a stable system where some **97** is still present. b. The reaction was changed to pH 12, driving the reaction to completion. Spectrum acquired 12 h after pH change showing complete conversion to **49**.*

Mechanistically, the stability of the intermediate species seems to indicate that the cyclisation does not take place through nucleophilic attack of the C1′-hydrate oxygen atom onto the epoxide moiety C4′, most likely because the strained tetrahedral geometry of C4′ does not allow for the linear S_N2 mechanism[307] to take place *via* 5-*exo*-*tet* cyclisation. Therefore, the mechanism occurs *via* addition of hydroxide ion to the least hindered carbon atom of the epoxide moiety C5′, leading to epoxide ring opening liberating a secondary hydroxyl moiety at C4′ which can undergo intramolecular addition to C1′ carbon atom, to form bicyclic aminooxazoline (**49**; *Scheme 4.4*, *vide supra*). This is consistent with the observation that high pH favours cyclisation.

The reaction of 2-aminooxazole (**55**) and glycidaldehyde (**69**) to form oxazolines **97** at pH 5 was found to be lower yielding than expected, considering it appeared nearly quantitative when initially analysed by ^{1}H NMR (*Figure 4.2*, *vide supra*). However, upon comparative integration against an internal standard present, we found that the yield of **97** was of 68%. To optimise the reaction conditions, extensive screens of concentration, temperature, pH, buffer, were undertaken. The highest yield of **97** obtained was in more dilute reactions (74% yield; 63mM **69**, 1.1 eq **55**, pH 5, RT). However, only a slight improvement on the initial reaction of equimolar concentrations of **69** and **55** (128mM; 68%) was observed. We hypothesise that the instability of epoxide **69** was the limiting factor, but despite the loss of material due to unknown side reactions, these do not seem to interfere with our chemistry or the analysis of these reactions (or future phosphorylation experiments, *vide infra*, *Section 4.3*.).

Interestingly, the aminooxazolines (**49**; *Figure 4.1*, *vide supra*) obtained in the hydrolysis of **97** provided a different stereoselectivity compared to the reaction of glyceraldehyde (**12**) with 2-aminooxazole (**55**). **12** reacts with **55** to give 44:34:13:9 *ribo*-/*arabino*-/*lyxo*-/*xylo*-**49** diastereomeric ratios (*Table 4.2*, entry 1). In order to analyse the correct ratio and selectivity of the four pentose aminooxazoline diastereomers in our reactions, we repeated the pH 5 reaction and monitored it until the system reached the same stable point as previously (12 h at

RT). In order to make sure that all the starting epoxide was consumed, so that there would be no competing chemistry between **69** and its hydrolysis product **12** (and subsequent reaction with **55**), we performed the reaction with a slight excess of **55** (1:1.3 ratio of **69**:**55**). Once clean synthesis of **97** was achieved, we separated the reaction mixture into several parallel reactions where the pH was raised by different amounts, in order to assess at which point the **97** oxazolines hydrolysed and cyclised to form **49**. The reaction best suited for the analysis of the diastereotopic selectivity was when the pH was raised to 12 as there was near total conversion of **97** to **49** after only 12 h (<7% **97** based on ^{1}H NMR integration of the 5.7 ppm **97** region and 6 ppm **49** region). Quantitative conversion to **49** was observed after 20 h (^{1}H NMR analysis with internal standard, with regards to **97** concentration), yielding **49** in 70% over two steps (*Table 4.1*, entry 2). ^{1}H NMR integration in this reaction showed a ratio of 42:16:16:26 *ribo*-/*arabino*-/*lyxo*-/*xylo*-**49** (based on ^{1}H NMR integration of the whole region around the pentose furanosyl doublets and the *lyxo*-**49**p doublet downfield, *Figure 4.6, Table 4.2*, entry 2).

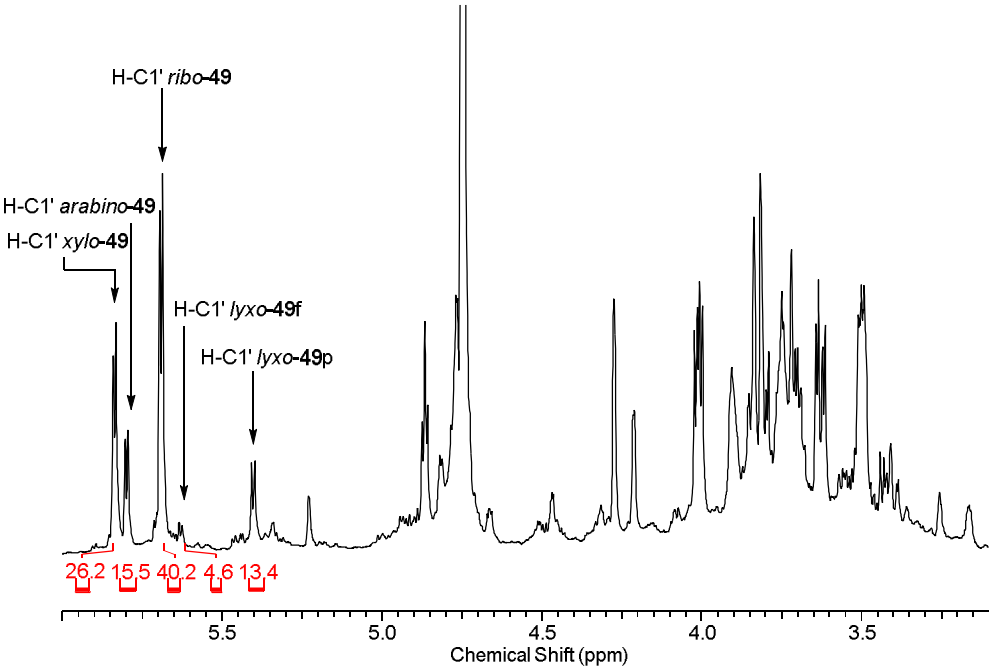


*Figure 4.6: Example of ratio calculation for the four stereoisomers of pentose aminooxazolines (**49**) by ^{1}H NMR spectrum integration (600 MHz, H_2O/D_2O 9:1, 6.00-3.20 ppm). Integration and assignment is shown for the H-C1' peaks of **49**. Ratio calculation is based on integration of the pentose furanosyl doublets and the lyxo-**49**p doublet downfield adding up to 100. The reaction was carried out with glycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.3 eq) in water at pH 5 and at RT. The pH was increased to 12 after 13 h. Spectrum acquired after 22 h at pH 12.*

Extensive screens of oxazolines **97** formation reactions were undertaken in order to see if the ratios of **97** and subsequent aminooxazolines ratios varied, by changing concentration, temperature, pH, buffer, ratios, and screening the effect of various metal ions (Mg^{2+}, Li^{+}, Ca^{2+}, Na^{+}). None of these reactions showed any significant difference in ratios (*Table 4.3*), suggesting the selectivity is inherent to the initial reaction of glycidaldehyde (**69**) and **55**.

Entry n°	Reaction	^{1}H NMR integration relative to other aminooxazolines integrated (%)				
		ribo	*arabino*	*lyxo* (f)	*lyxo* (p)	*xylo*
1	**69** + **55** + Ca	43	14	4	13	26
2	**69** + **55** + Li	45	17	4	10	24
3	**69** + **55** + Mg	44	14	2	12	28
3	**69** + **55** + Na	44	17	3	11	25

*Table 4.3: Stereoselectivity in the reaction of glycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.3 eq) in the presence of metal salts (2 eq) at pH 5, RT, followed by hydrolysis of adduct **97** by addition of hydroxide. The ratio calculation for the four stereoisomers of aminooxazoline (**49**) was performed by ^{1}H NMR spectra integration (600 MHz, H_2O/D_2O, 9:1) as a percentage of the H-C1' peaks.*

Equimolar reactions of 2-aminooxazole (**55**) and glycidaldehyde (**69**) were set-up at RT, at a range of pH, from 3 to 12, in order to test the relationship between diastereomeric ratios and initial pH. At pH 3, **69** is stable and not observed to react with **55**. This is most likely due to the protonation of **55** (pK_a 5.4), making it less nucleophilic and therefore less prone to attacking **69**[308]. After 30 h the reaction had yielded <20% **97** (relative percentage) and there is significant competing hydrolysis of **69** to **12** (29%, based on ^{1}H NMR integration of the 5.7 ppm **97** region and **12** and **69**). Interestingly, at pH 9 and 12 however, **55** and **69** react to give aminooxazolines (**49**) after 24 h and 5 h respectively. At pH 12, the reaction is very fast and **49** are observed after 1 h, and there is no residual **69** after 5 h (15% yield **49** by ^{1}H NMR analysis with internal standard, *Table 4.1*, entry 3). Interestingly there is a similar selectivity of *ribo->xylo->>arabino->lyxo*-**49** (*Figure 4.7; Table 4.2*, entry 3) to that found in the previous two step reaction (where the oxazolines **97** are formed at pH 5 and hydrolysed to **49** at pH 12, as depicted in *Scheme 4.4*, *vide supra*). This suggests that **49** formation occurs transiently *via* the oxazolines **97**, rather than by hydrolysis of **69** to **12** and subsequent reaction with **55**.

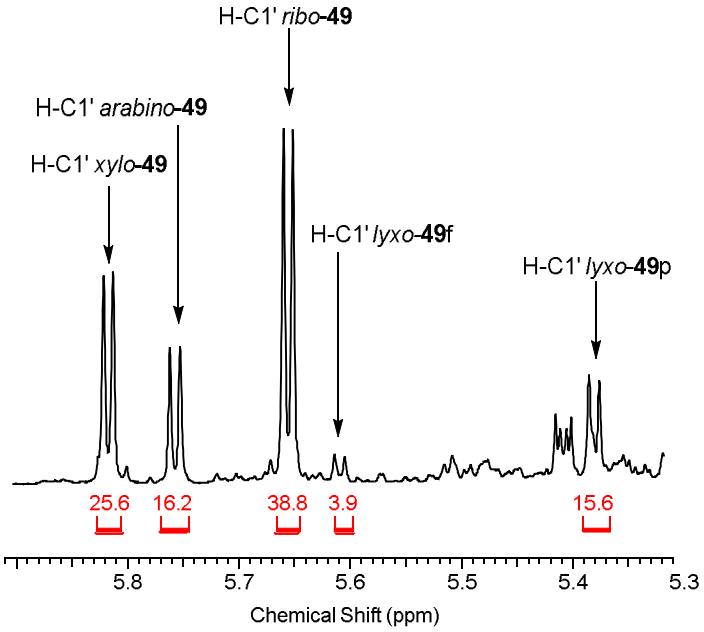


*Figure 4.7: ^{1}H NMR spectrum integration (600 MHz, H_2O/D_2O 9:1, 5.91-5.30 ppm). Integration and assignment is shown for the H-C1′ peaks of aminooxazolines (**49**). Ratio calculation is based on the percentage of the integration of the pentose furanosyl **49** doublets and the lyxo-**49**p doublet downfield. The reaction was carried out with equimolar concentrations of glycidaldehyde (**69**) and 2-aminooxazole (**55**) in water at pH 12. ^{1}H NMR acquired after 8 h.*

There is a significant difference in diastereomer selectivity between these glycidaldehyde (**69**) reactions and the previously reported glyceraldehyde (**12**) reaction (*Table 4.2*, entry 1). Selectivity seems to be governed by the initial attack of **69** and **55**, forming oxazolines **97**. We observed four distinct **97** species in solution (*vide supra*, *Figure 4.2*), which, when integrated, seem to correlate with the observed diastereomer selectivity of the **49** after hydrolysis of **97** (*Table 4.2*, entries 7 and 2 respectively[x]). However, there is bias towards *ribo*- and *xylo*-**49** due to possible C2′ epimerisation from *arabino*- to *ribo*- and *lyxo*- to *xylo*- respectively[309] (a point to be returned to later, *Section 6*). This bias can be

[x] The four distinct doublets interpreted as H-C1′ (*Figure 4.2*, insert) were assigned as the compounds *ribo*-, *lyxo*-, *xylo*- and *arabino*-**49**/**53**, from left to right. This is due to similarities in integration, coupling patterns and the aforementioned stability of the *ribo*-leading **97**, enabling us to assign it confidently.

alleviated when looking at combined ratios of [*3R,4R*]-**49** (*ribo*- and *arabino*-) and [*3S,4R*]-**49** (*lyxo*- and *xylo*-) respectively (*Table 4.4*, entry 4).

Entry n°	Reaction	^{1}H NMR integration relative to other aminooxazolines integrated (%)		
		[*3R,4R*]-isomers	[*3S,4R*]-isomers	|Δ|
1	**12** + **55**	78	22	**-25**
2	**97** + OH^-	58	42	-5
3	**69** + **55** to **49**	55	45	-2
4	**68** + **55**	86	14	**-33**
5	**97** + P_i	61	39	-8
6	**69** + **55** + P_i	56	44	-3
7	**69** + **55** to **97**	53	47	0
8	**102** + **55**	19	81	**34**
9	**69** + **55** + Ca	57	43	-4
10	**69** + **55** + Li	62	38	-9
11	**69** + **55** + Mg	58	42	-5
12	**69** + **55** + Na	61	39	-8

*Table 4.4: Stereoselectivity in various aminooxazoline assembly reactions. Ratios are a result of addition of the ribo- and arabino- ratios, and the xylo- and lyxo- ratios ([3R,4R]-isomers and [3R,4R]-isomers, respectively) found in Table 4.2 (for entries 1-8) and Table 4.3 (for entries 9-12). Reactions described in Scheme 4.6. |Δ|: The absolute difference between the combined [3R,4R]-isomers ratio of a reaction and the corresponding ratio of **97** formed from the reaction of **69** and **55** at pH 5, entry 7 (blue). Differences of more than ±10 are marked in bold.*

In the reaction of glyceraldehyde (**12**) and 2-aminooxazole (**55**) it was rationalised that the selectivity observed was due to preferential attack of the *Si* face of the aldehyde, due to an intramolecular hydrogen bond (C3-O-**H**–**O**=C1) locking the conformation of the aldehyde and lowering the transition state energy of the reaction[272] (*Figure 4.8*). This hydrogen bond has also been used to explain the selectivity observed in the reaction of a glyceraldehyde imine **102** (formed from the reaction of **12** with AICA **21** or AICN **20**) with **55**, leading to a reversal in the observed facial selectivity of **12** where the *lyxo*-conformer *lyxo*-**58** was the main compound formed[274] (*Table 4.2*, entry 8; a point to be returned to later, *Section 7*).

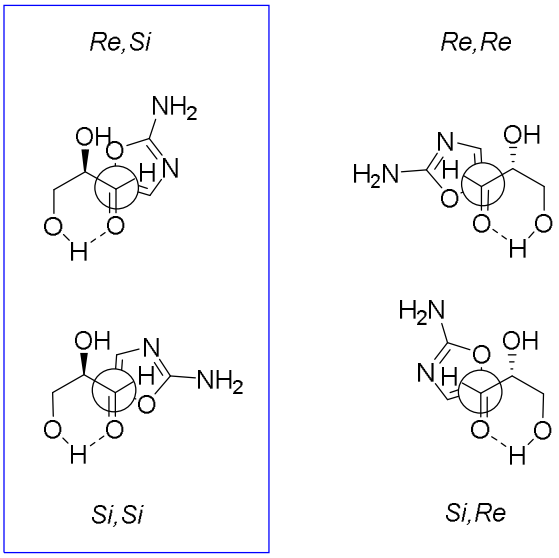


Figure 4.8: Rationale for the stereoselectivity observed, favouring ribo- and arabino- (blue box: top and bottom, respectively). The stereochemical descriptions[xi] *refer to 2-aminooxazole (**55**) and C=O of glyceraldehyde (**12**) respectively. The structures are shown in the idealised staggered transition structures (adapted from the SI of Anastasi et al.*[272]*).*

However, the epoxide moiety of **69** blocks the potential to adopt the H-bond motif, thereby losing selectivity for attack at the *Si* face of **69**. However, we now observe preferential attack from one face of the nucleophile **55** (*Re*; *Figure 4.9*). We hypothesise this could be due to a change in the transition state energy. Indeed, the hydrogen bond proposed in the **12**-**55** reaction (*Figure 4.8*) would lower the energy needed to attack the carbonyl. Any difference in energy exhibited by one face of the nucleophile over the other might have been lost due to the lower overall energy of the system. The slight preference for the *ribo*- over the *arabino*- variant in **12** reaction (44%:34% respectively, *Table 4.2*, entry 1) might be reflective of this inherent preference in reactivity of one face (*Re*) of the nucleophile. It is possible that this small difference between *ribo*-/*arabino*-stereochemistry observed during the **12** and **55** reaction is magnified in the **69** and **55** reactions upon losing the C3-OH intramolecular general acid catalyst. Blocking the C3-OH moiety would remove the potential for hydrogen bonding to

[xi] Designed *Re* if the groups of the trigonal atom described are clockwise, or *Si* if anti-clockwise, using Cahn-Ingold-Prelog priority rules. [xi]

lower the energy of the corresponding transition state. Accordingly, the inherent difference from the two faces of the nucleophile would be more prominent in the **69** reactions due to a large ground state to transition state energy gap, leading to a more selective attack from the *Re* face of the nucleophile. This in turn leads to a stereoselective formation of the **97** variants leading to [1*S*,2*R*,4*R*]-**49** (*ribo*- and *xylo*-) over [1*R*,2*S*,4*R*]-**49** (*arabino*- and *lyxo*-).

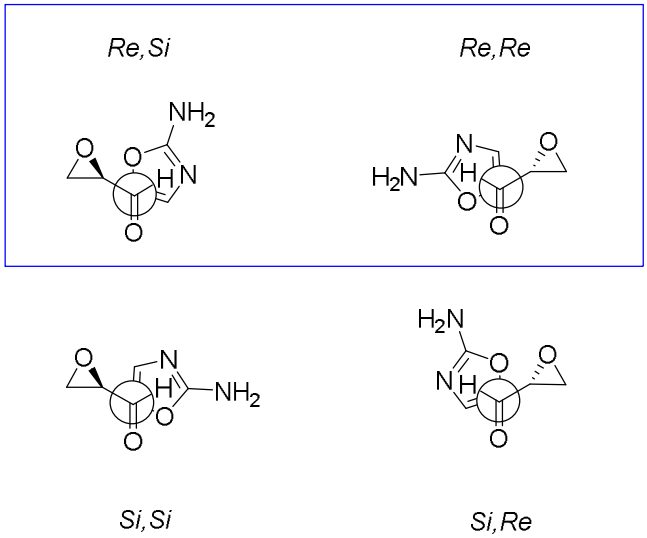
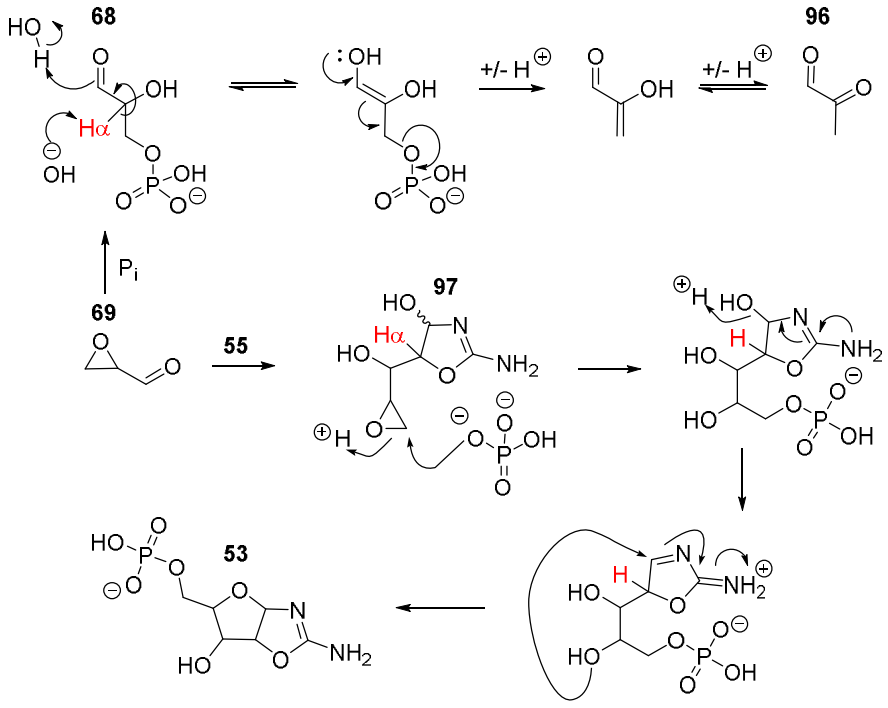


Figure 4.9: Stereoselectivity observed, favouring ribo- and xylo- (blue box: left and right, respectively). The stereochemical descriptions[xi] *refer to 2-aminooxazole (**55**) and C=O of glycidaldehyde (**69**) respectively. The structures are shown in the idealised staggered transition structures. Only compounds leading to D isomers are depicted.*

The observed formation of the **97** adduct shows it is possible to undertake an aldol reaction with the aldehyde moiety of **69** while preserving the epoxide moiety and its reactivity intact. This could help solve the phosphorylation problem mentioned earlier (*Section 4.1*), whereas in the event of a nucleophilic phosphorylation at the epoxide C3, the aldehyde portion of **69** initiates the process leading to the elimination of phosphate via E1cB mechanism (*vide supra*, *Scheme 4.3*). We hypothesised a new nucleophilic phosphorylation pathway, without the risk of elimination, through initial aldol reaction of **69** and **55** to yield oxazolines **97** and subsequent terminal phosphorylation of the epoxide moiety, bringing us closer to an aqueous phosphorylation of nucleotides.

4.3. Aqueous phosphorylation of aminooxazolines

The stability of oxazolines **97** at pH 5 gave us an ideal window of opportunity for nucleophilic attack of the epoxide moiety, possibly leading to 5′-modified aminooxazolines. We decided to investigate the synthesis of 5′-phosphorylated aminooxazolines (**53**) utilising the stability of adduct **97** and the strained nature of the epoxide ring to direct regiospecific nucleophilic 5′-phosphorylation (*Scheme 4.7)*. Previously, we had observed the selective and efficient phosphorylation of glycidaldehyde (**69**) with P_i to form glyceraldehyde-3-phosphate (**68**), but subsequent elimination of phosphate gave methylglyoxal (**96**) and accordingly resulted in low yields of desired phosphorylated sugar **68** (*vide supra*, *Section 3*). However, phosphorylation of **97** would yield a δ-substituted phosphate, rather than the β-substituted phosphate of **68**. The distal position of the δ-phosphate, with respect to the aldehydic/anomeric centre was predicted to completely supress detrimental E1cB-type elimination.



*Scheme 4.7: Proposed mechanism for the reaction of glycidaldehyde (**69**) with P_i (top) forming glyceraldehyde-3-phosphate (**68**) followed by E1cB elimination to methylglyoxal (**96**) (top); and the reaction of **69** with 2-aminooxazole (**55**) to form adduct **97**, and subsequent addition of P_i to form 5'-phosphorylated furanosyl-aminooxazolines (**53**) (bottom). Hα is shown in red to highlight the distal position of the phosphate once added to **97**.*

In order to simplify analysis and selectively study the insertion of phosphate in **97** (without any competing chemistry), we reacted glycidaldehyde (**69**) and 2-aminooxazole (**55**) at pH 5 to form oxazolines **97** (as described above, *Section 4.2.*), and used separate aliquots of that reaction to study the effect of different concentrations of P_i, at different pHs, on nucleophilic insertion at the epoxide moiety of **97**, leading to aminooxazoline formation. Similarly to the reactions done without phosphate present (*vide supra*), one of the **97** intermediates was slower to convert. However, very encouragingly, it was now observed to undergo complete conversion to aminooxazolines at neutral pH (*Figure 4.12*, *a.*; *vide infra*). This can be explained by the higher nucleophilicity of phosphate at pH 7 (pK_as of phosphate - pK_{a1} 2.1, pK_{a2} 7.2 and pK_{a3} 12.6), compared with water (pK_a 14).

Extensive analysis was carried out on the reaction of **97** with an excess of P_i (3 to 8 eq) at neutral pH, in order to assess whether the aminooxazoline species formed were 5′-phosphorylated. ^{31}P NMR analysis shows clear coupling of phosphate to the 5′-region of the aminooxazolines present (*Figure 4.10*, *a*). Selective TOCSY NMR analysis of one of the major compounds in solution (H-C1′ shift: 5.8 ppm) displays clear evidence of phosphorus coupling (*Figure 4.10*, *b)*. This was interpreted to be riboaminooxazoline 5'-phosphate (*ribo*-**53**) due to the similarities in proton shifts and coupling patterns to riboaminooxazoline (*ribo*-**49**; H-C2', t, *J* = 5.3 Hz). Further evidence that we are mainly forming **53** is the lack of any visible *lyxo*-**53** peaks. In the event of *lyxo*-**53** formation, we recognised that we would not observe the H-C1′ doublet corresponding to the pyranosyl form of the aminooxazoline. Steric hindrance, due to all the constituents being on the same side (including the bulky and electron-dense phosphate group), was anticipated to limit furanosyl ring closure (*Figure 4.11*), and pyranosyl ring formation (which is the favoured isomer in the *lyxo*-**49** variant) would be blocked by the presence of the 5′-phosphate. Instead the *lyxo*-compound would preferentially sit as the open chain aromatic species **100**a or anomerise to the *xylo*-**53** (a point to be returned to later, *Section 6*).

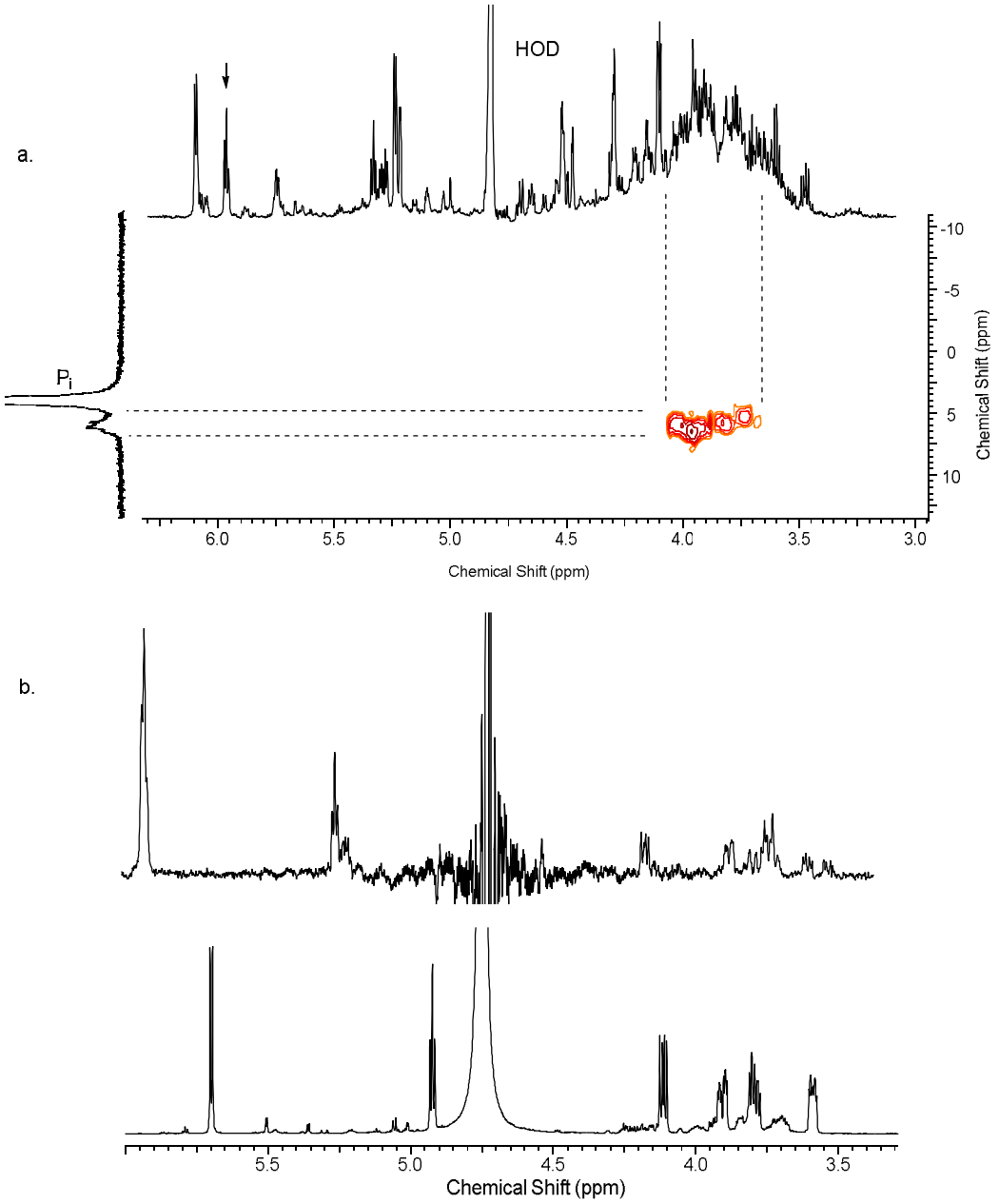


*Figure 4.10: a. ^{1}H-^{31}P HMBC (400 MHz, H_2O/D_2O 9:1) showing coupling between phosphorus and protons in the 3.5-4 ppm region, evidence of successful phosphorylation at the C5' position of aminooxazolines (**53**). b. ^{1}H NMR spectrum (600 MHz, H_2O/D_2O 9:1, 6.00-3.30 ppm) comparing a selective-TOCSY[xii] spectra (top) of H-C1' peak annotated with arrow (in a.) and ^{1}H NMR of synthetically prepared ribo-**53** (bottom). H-C1' and H-C2' shift due to pH difference.*

[xii] Selective-TOCSY allows for the isolation of the signal for protons within the same coupling network.

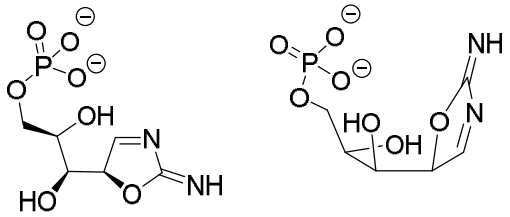


*Figure 4.11: The lyxo-stereoisomer of the open chain aminooxazoline 5'-phosphate (**53**) before cyclisation. The bulky phosphate group on the C5' leads to steric clash with the C4'-OH and the 1-2 oxazoline ring, disfavouring ring closure.*

In order to unequivocally characterise the major species as **53**, we saturated the solution in P_i in order to drive the synthesis towards **53**. 8 equivalents of P_i (with regards to original concentration of epoxide **69**) were added to oxazolines **97** and the pH was adjusted to 7, in order to push the reaction selectively towards phosphorylation while limiting competing hydrolysis (we previously observed **97**'s propensity to rapidly hydrolyse to **49** at pH 9 but not pH 7, *vide supra*). After 11 days the reaction was deemed to have gone to completion as no **97** was left in solution. However, the open chain hydrate **101** and aromatic form **100** of the aminooxazolines were present (characteristic **101** doublets around 5.5 ppm - 5.49, d, J = 2.5 Hz; 5.48, d, J = 2.4 Hz; and **100** singlets – 6.60 ppm, s; 6.58 ppm, s; respectively; a point to be returned to later, *Section 6*). In order to simplify analysis, we sought to convert those aromatic and open chain species back to their furanosyl-ring counter parts by raising the pH. After a further 5 days at pH 9, these peaks decreased to yield a simpler NMR with 3 major aminooxazoline species (*Figure 4.12*).

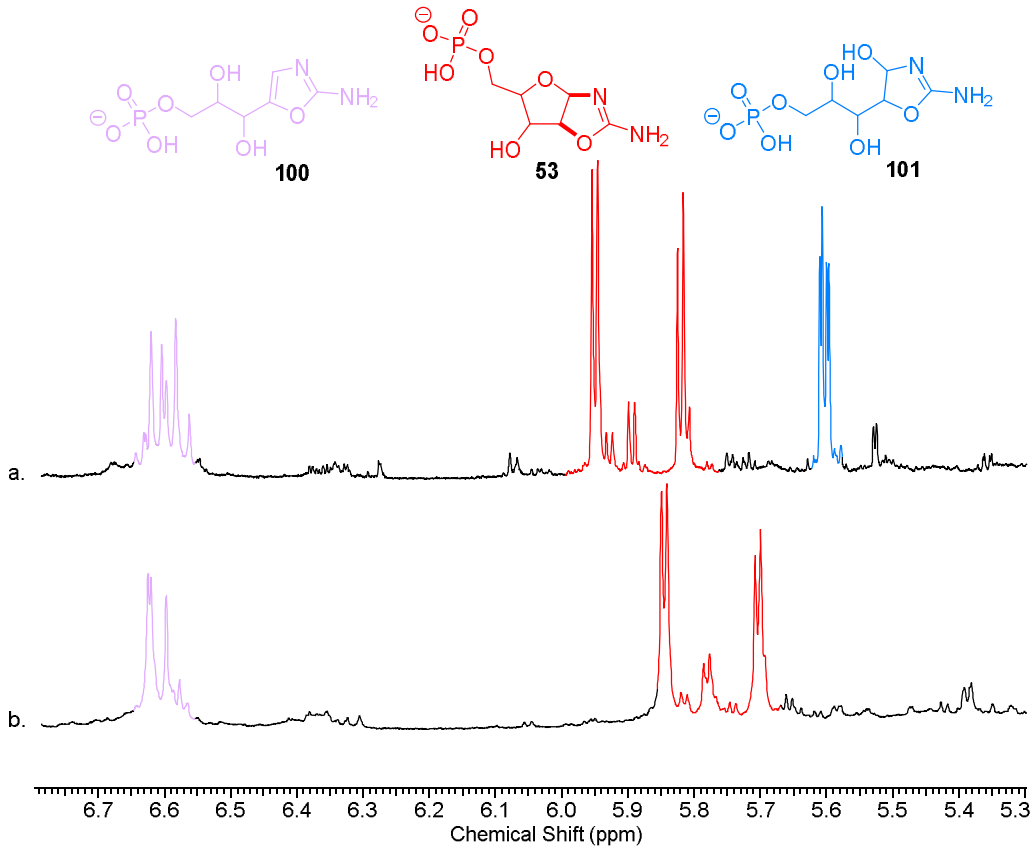


*Figure 4.12: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 6.80-5.30 ppm) showing the H-C1' peak region for aminooxazolines, and the aromatic region. a. ^{1}H NMR showing the formation of aminooxazolines 5'-phosphate (**53**) after 11 days at pH 7, along with the characteristic peaks of the open-chain hydrates **101** and open-chain aromatic **100** at 5.6 ppm and 6.6 ppm respectively. b. ^{1}H NMR showing conversion of those peaks to **53** after a further 5 days at pH 9.*

Spiking with synthetically prepared samples of *ribo*-, *xylo*- and *arabino*-**53** (*Figure 4.13*) confirmed the presence of 5′-phosphorylated aminooxazolines (**53**) as the major species (30% yield over 2 steps, based on initial epoxide **69** concentration by ^{1}H NMR spectra analysis with internal standard, *Table 4.1*, entry 5). ^{31}P NMR shows increase of the peaks attributed to **53** with no new phosphorus peaks (*Figure 4.14*), further confirming correct interpretation of our spiking. Spiking with pre-incubated open-chain *lyxo*-**53** confirms the presence of the *lyxo*- stereoisomer sitting in its aromatic form **100**a (characteristic singlet at 6.7 ppm, *Figure 4.13, d.*), as previously speculated. Lastly, spiking with synthetically

prepared samples of *ribo*-**49** and *xylo*-**49** shows an inconclusive spike with the *xylo*-stereoisomer but a conclusively negative spike for *ribo*-**49**, suggesting the majority of the aminooxazolines observed are 5'-phosphorylated (*Figure 4.15*).

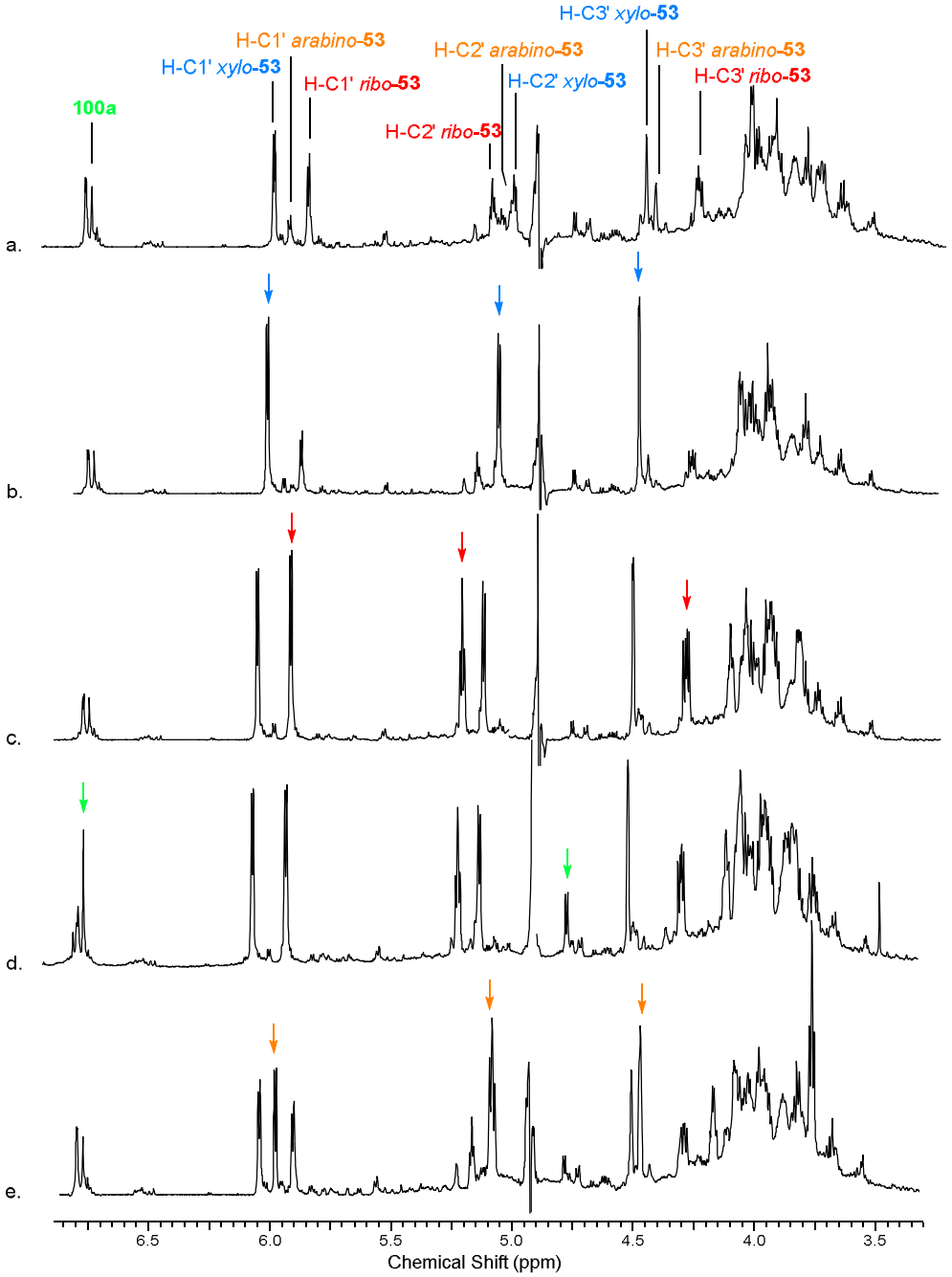


*Figure 4.13: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 6.85-3.30 ppm) showing signal amplification observed by spiking with synthetically prepared authentic standards of aminooxazolines 5'-phosphate (**53**). a. Reaction of oxazoline **97** with 8 equivalents of P_i at pH 7 with subsequent increase to pH 9. Spiked with b. xylo-**53** c. followed by ribo-**53** d. and pre-incubated lyxo-**53**. e. Another aliquot of crude reaction mixture (a.) spiked with arabino-**53**.*

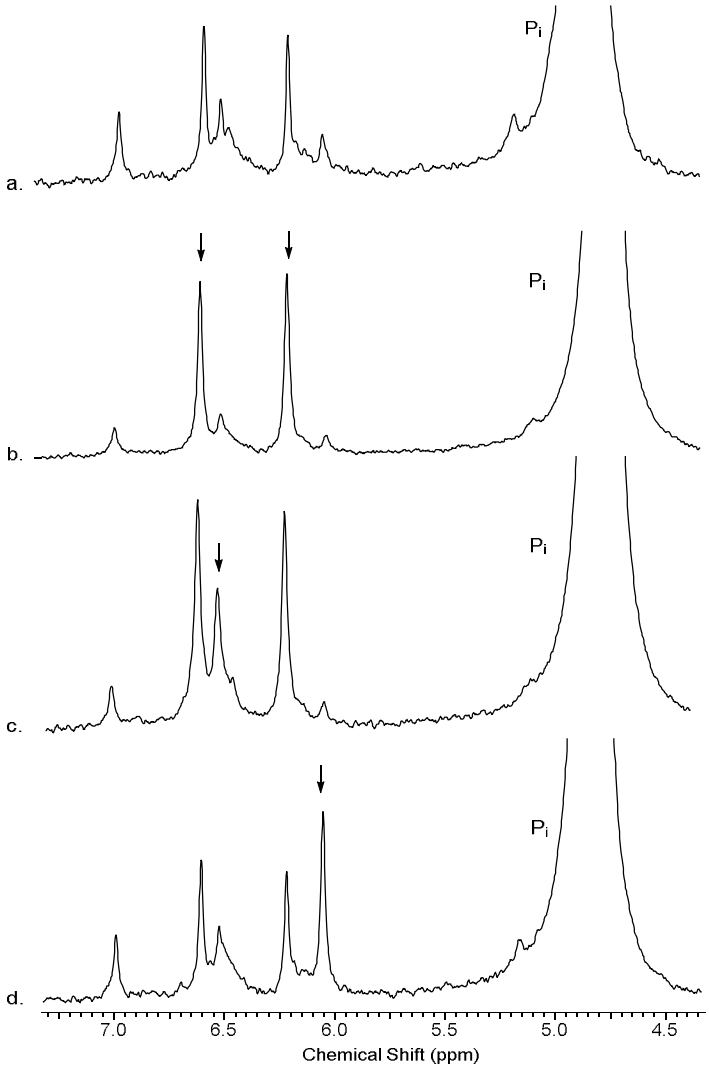


*Figure 4.14: ^{31}P NMR spectra (300 MHz, H_2O/D_2O 9:1, 7.30-4.35 ppm) showing signal amplification observed by spiking with synthetically prepared authentic standards of aminooxazolines 5'-phosphate (**53**). a. Reaction of oxazoline **97** with 8 equivalents of P_i at pH 7 with subsequent increase to pH 9. Spiked with b. ribo-**53** and xylo-**53** c. followed by pre-incubated lyxo-**53** d. Another aliquot of crude reaction mixture (a.) spiked with arabino-**53**.*

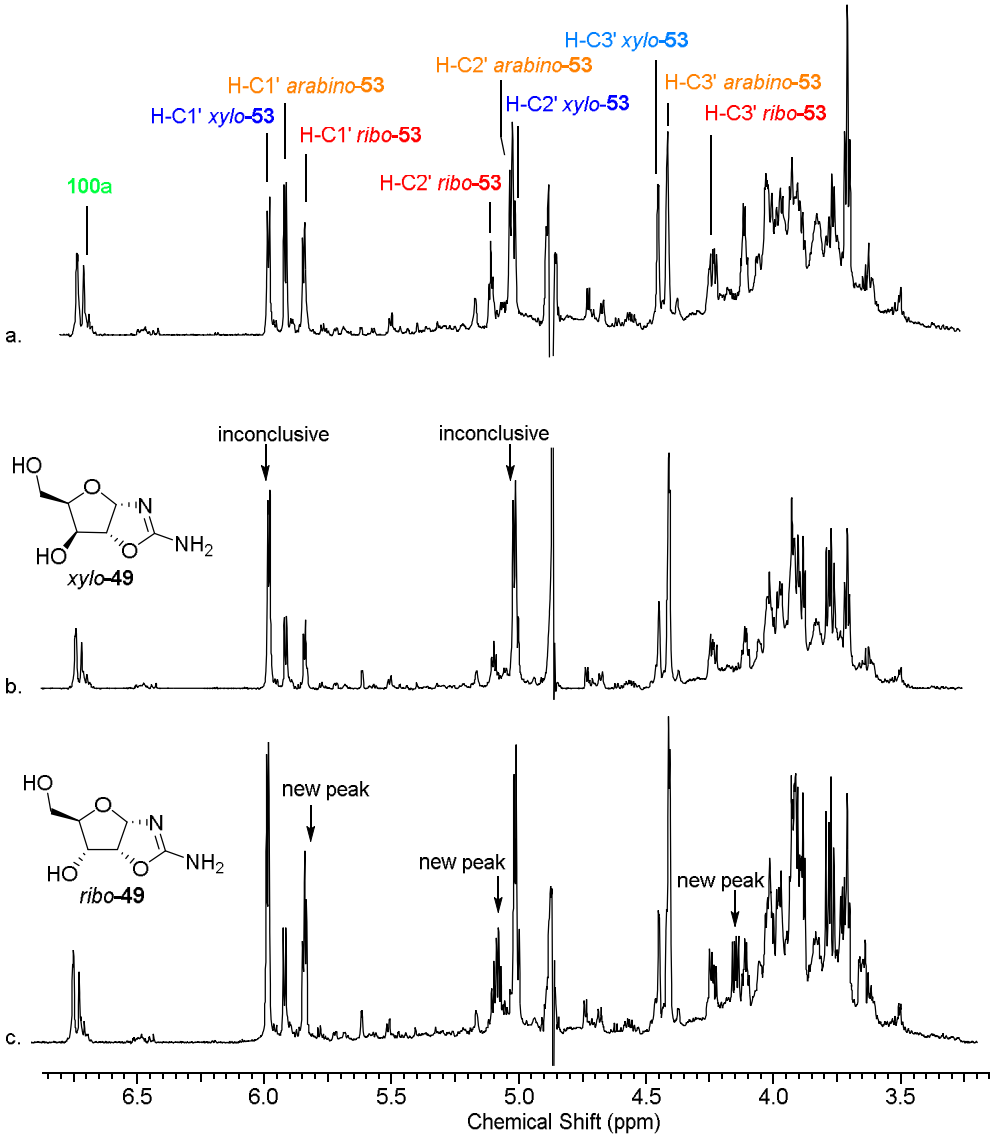


*Figure 4.15: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 6.85-3.15 ppm) showing signal amplification observed by spiking with synthetically prepared authentic standards of aminooxazolines 5'-phosphate (**53**). a. Reaction of oxazoline **97** with 8 equivalents of P_i at pH 7 with subsequent increase to pH 9 and spiked with arabino-**53** (Figure 4.13). Spiked with aminooxazolines (**49**) b. xylo-**49** c. followed by ribo-**49**.*

Spiking a reaction with less P_i present (3 eq P_i) with synthetic sample of *ribo*-**53** provided correct amplification of the expected peaks, showing that even without an overwhelming excess of P_i, **53** are the main species (Appendix - *Figure 11.3*). All evidence suggested that phosphorylation was successfully achieved at the

5'-position of **97**, showing that our proposed mechanism accurately represents the chemistry at play (*Scheme 4.7*, *vide supra*). The phosphorylation was successful, and there seemed to be only a minor amount of aminooxazolines **49** by-products (minor H-C1′ *ribo*-**49** peaks can be detected in the ^{1}H NMR spectra, Appendix - *Figure 11.3*, *c.*).

In order to analyse the ratio of the phosphorylated isomers **53** without any interference from trace amounts of un-phosphorylated **49**, the above reaction was passed through a hydroxide ion exchange (Dowex®) column in order to retain the phosphorylated species selectively, which were then eluted using formic acid solutions of increasing concentration. The resulting formic acid fractions were lyophilised and the lyophilite dissolved in D_2O for analysis. ^{1}H NMR integration in the reaction of oxazolines **97** with an excess of P_i to form 5'-phosphorylated aminooxazolines **53**, showed a ratio of 45:16:0:39 *ribo*-/*arabino*-/*lyxo*-/*xylo*-**53** (based on ^{1}H NMR integration of the three stereoisomers H-C1′ peaks around 6.2 ppm, and reported as an average of percentages over three NMRs, *Table 4.2*, entry 5). Selectivity of the stereoisomers of **53** follows the same order as alkaline synthesis of **49**, *ribo*>*xylo*>>*arabino*>*lyxo* (*Figure 4.16*). There is however a more pronounced preference for the *xylo*-stereoisomer in **53**, which represents 39% of **53** *vs* 26% of **49**. This difference is most likely linked to the observation that *lyxo*-**53** is sitting exclusively in the aromatic open-form **100**a, a compound that is planar at C2′ and therefore allows for C2′-epimerisation (a point to be returned to later, *Section 6*), which would increase the proportion of *xylo*-**53**. However, the combined ratios of [3*R*,4*R*]-**53** (*ribo*- and *arabino*-) *vs* [3*S*,4*R*]-**53** (*lyxo*- and *xylo*-) (*Table 4.4*, entry 5), obviating any changes in ratios due to C2′-epimerisation, are observed to be almost identical to the high pH hydrolysis of **97** to **49** (*Table 4.4*, entries 7 and 2 respectively), demonstrating that the stereoselectivity is governed by the initial attack of 2-aminooxazole (**55**) on glycidaldehyde (**69**).

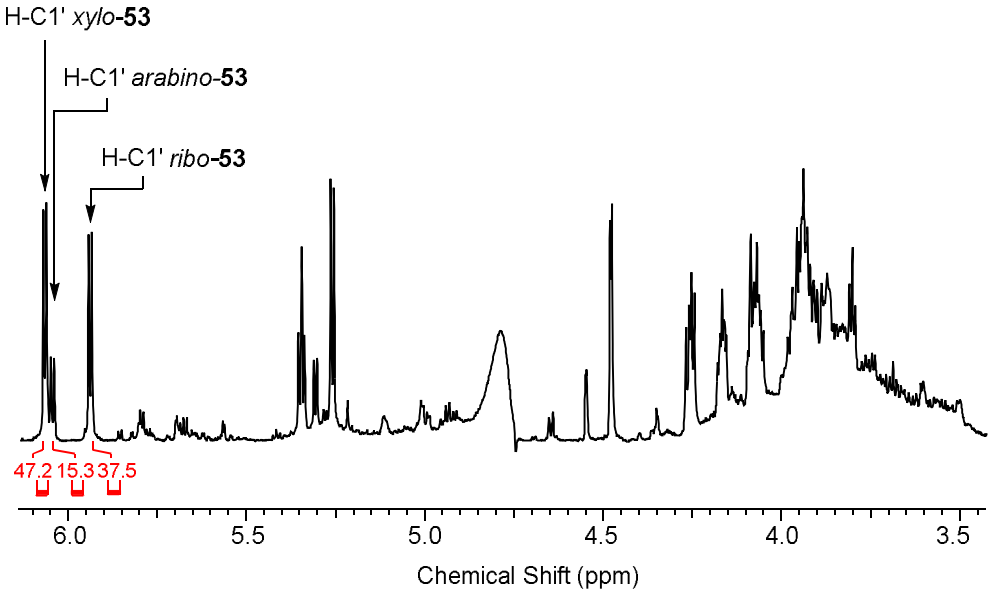


*Figure 4.16: Example of ratio calculation for the three stereoisomers of aminooxazolines 5'-phosphate (**53**) by ^{1}H NMR spectrum integration (600 MHz, D_2O, 6.10-3.45 ppm). Integration and assignment is shown for the H-C1' peaks of **53**. Ratio calculation is based on integration of the three stereoisomers H-C1' peaks around 6 ppm adding up to 100. Reaction of oxazoline **97** with 6 equivalents of P_i at pH 7 for 14 days, after which the phosphorylated species were isolated by OH^- Dowex® column and subsequently eluted with formic acid, lyophilised and diluted in D_2O for analysis.*

We were interested to see if the reaction occurred similarly and yielded a comparable diastereotopic distribution to the above reactions if P_i was present from the start. Reactions of glycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.1 eq) set-up in P_i buffer at a range of pH (5 to 9), by-passing the intermediate **97** stage, also yielded 5'-phosphorylated aminooxazolines (**53**) (and minor amounts of non-phosphorylated **49**; 21% combined, *Table 4.1*, entry 6; Appendix - *Figure 11.4*). Again, analysis of the diastereomer ratio was undertaken at neutral pH, after separating the phosphorylated species from the crude reaction by a hydroxide Dowex® ion-exchange column, and subsequent elution with aqueous formic acid. The ratio obtained was similar to the one above, where *ribo*-**53**>*xylo*-**53**>>*arabino*-**53** (*Table 4.2*, entry 6), suggesting the diastereomer distribution was not affected by the initial presence of P_i. We decided to reexamine the reaction reported by Anastasi *et al.* where they formed **53** from

glyceraldehyde-3-phosphate (**68**) and 2-aminooxazole (**55**)[283]. Their analysis of the **53** ratios was performed post-phosphatase treatment[283]. In order to compare the results with our **53** yielding reactions involving epoxide **69**, we repeated the reaction with 1:1 **68**:**55** (pH 7, 60mM, 4 days), and analysed the results similarly to our reactions of glyceraldehyde (**12**) and **55**, and epoxide **69** and **55** (*Table 4.2*, entry 4 and entries 3,5,6 respectively), by integrating the **53**/**49** region by ^{1}H NMR. The diastereomers ratio trend observed was *ribo*-**53**>>*arabino*-**53**>*xylo*-**53** (25% **53**, *Table 4.1*, entry 4), similarly to the **12** and **55** reaction. No cyclised *lyxo*-**53** was observed. Interestingly, Anastasi *et al.* report 11% *lyxo*-**49** post-phosphatase treatment, suggesting that while some *lyxo*-**53** was most likely formed in our repeat of the **68** and **55** reaction, *lyxo*-**53** was not cyclised in its aminooxazoline form and thus *lyxo*-**53** was not taken into account in our ratio calculations[xiii].

The significant difference in stereoselectivity between this reaction of glyceraldehyde-3-phosphate (**68**) and 2-aminooxazole (**55**) (where *xylo*-**53** is a minor product) and the reaction of glycidaldehyde (**69**), **55** and P_i (where *xylo*-**53** is a major product alongside *ribo*-**53**; *vide supra*), suggests that **53** formation in our three component reaction occured *via* the transient adduct **97** (*Scheme 4.7*), rather than by phosphorylation of epoxide **69** to **68** and subsequent reaction with **55** (also see *Section 4.2*.). This further reinforced the hypothesis that the diastereomer selectivity is fixed during the initial addition of **55** to **69** to form adduct **97**, and any future reaction will reflect that initial stereoselectivity.

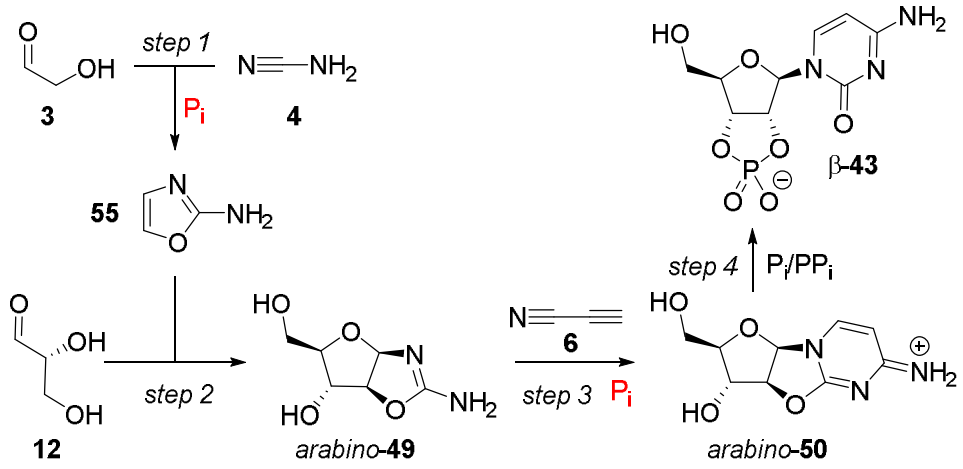
These results demonstrate the aldehydic position of glycidaldehyde (**69**) can react *via* aldol to form oxazolines **97** (transiently, pH 7; or stably, pH 5), while retaining the reactivity inherent to the epoxide intact, delivering selective and effective nucleophilic attack of phosphate at the desired position. Accordingly, 5′-phosphate aminooxazolines (**53**) are obtained *via* a novel, prebiotically plausible pathway, while solving the phosphate-elimination problem inherent to **68**.

[xiii] For consistency, any aromatic peaks representing the open-chain **53** were not integrated (in a similar manner to the experiment of **97** and phosphate).

Furthermore, in contrast with previous prebiotic phosphorylation attempts using a heated, dry-state[33,225] or a non-aqueous solvent[226], these results demonstrate a mild, and entirely aqueous phosphorylation, which works in relatively low concentrations of phosphate (400mM) at neutral pH and without heating. This is consistent with the bulk of the body of work pertaining to prebiotic syntheses, including all the work aforementioned. It is worth pointing out once again that the ultimate goal of any prebiotic research is to find a concomitant synthesis of not only all four ribonucleotides, but also a concomitant synthesis of all molecules important in biology or in its inception. Therefore, finding ways to bring us closer to a truly 'one-pot' synthesis is encouraged and necessary, which we have demonstrated here.

5. Multicomponent synthesis of 5′-aminooxazolines

Although the new prebiotic synthesis of pyrimidine nucleotides developed by the Sutherland group[33] helped solve the decades-old glycosidation problem that was central to the research into the origin of nucleotides (as detailed in *Section 4*, and depicted in *Scheme 5.1*), one fundamental problem in their new synthesis, as recognised by the authors, is the need for a step-wise addition of two and three carbon aldehydes to the reaction mixture in the initial step[167]. Robertson and Joyce also said "the story remains incomplete because these syntheses still require temporally separated reactions using high concentrations of just the right reactants, and would be disrupted by the presence of other closely related compounds"[167]. Indeed, step-wise addition imposes more constraints upon the prebiotic environment necessary for the reaction to occur, making the desired reaction less likely.

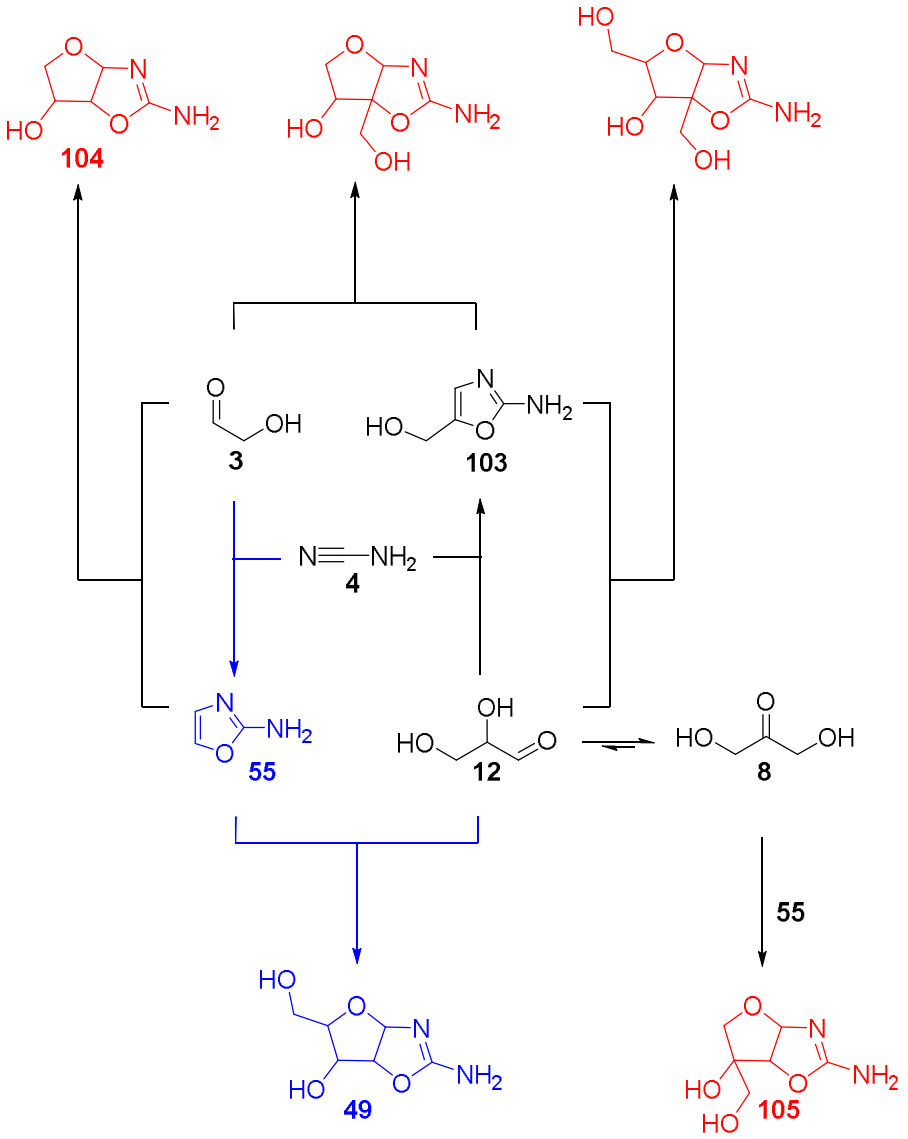


Scheme 5.1: Four step prebiotic synthesis of activated pyrimidine ribonucleotides as described by Powner et al[33]*, showing the sequential addition of C_2 and C_3 aldehydes glycolaldehyde (**3**) and glyceraldehyde (**12**), respectively. The intermediates shown depict those leading to the natural stereochemistry. Steps catalysed and/or buffered by P_i are shown in red.*

One central compound, formed early in the synthesis, is the sugar and base synthon 2-aminooxazole (**55**), formed from the reaction of C_2 aldehyde glycolaldehyde (**3**) and cyanamide (**4**). Although **4** and **3** can react in water to furnish **55** without the presence of P_i, this reaction requires strong alkaline conditions and heating[310]. In these conditions, glyceraldehyde (**12**), the C_3 sugar needed for the next step of the reaction, is unstable due to extensive aldol chemistry and isomerisation to its ketose form dihydroxyacetone (**8**)[297]. The Sutherland group found that these harsh conditions could be avoided by using P_i as a general acid/base catalyst[33]. P_i was used because phosphate forms an essential part of nucleotides (see β-**43**, *Scheme 5.1*), and thus would need to be introduced at some stage of the assembly. Furthermore, the second pK_a of P_i is close to neutral (pK_{a2} 7.2), allowing the reaction to be catalysed under mild conditions. Using P_i in the first step of the synthesis reaffirms the utility of P_i in prebiotic chemistry as catalyst (first step, *Scheme 5.1*), buffer (third step, *Scheme 5.1*) and reactant (fourth step, *Scheme 5.1*). High yields (>80%) of **55** could now be achieved prebiotically (from **4** and **3**) by using P_i as a catalyst at neutral pH, thus bypassing the need for elevated temperatures or strongly alkaline conditions.

Even though the Sutherland group found a solution to **55** synthesis, they found that the presence of **12** is incompatible with the initial reaction of **3** and **4** to form **55**. This is due to the fact that, in a mixture of the C_2 and C_3 aldehydes **12** and **3** at the start of the reaction, both could react with **4** indiscriminately to form oxazoles: (2-aminooxazol-5-yl)methanol (**103**) and 2-aminooxazole (**55**), respectively (*Scheme 5.2*). While **55** is the desired product of this reaction, the formation of **103** presents two problems: a depletion of the essential C_3 sugar stock **12** needed for subsequent reactions; and depletion of cyanamide (**4**), leading to diminished **55** formation, which in turn leads to left-over **3** (in the event of an equimolar reaction of **12** and **3**). Ultimately this leads to competition between the desired reaction of **12** and **55** (leading to pentose aminooxazolines; **49**), and all other reactions than can occur, due to the homologous nature of **12** and **3**, and **55** and **103** respectively (*Scheme 5.2*). Finally, this yields a complex mixture of products

from unwanted masked-aldol reactions, causing a problem of selection for canonical nucleotides via the desired product **49** (*Scheme 5.2*).

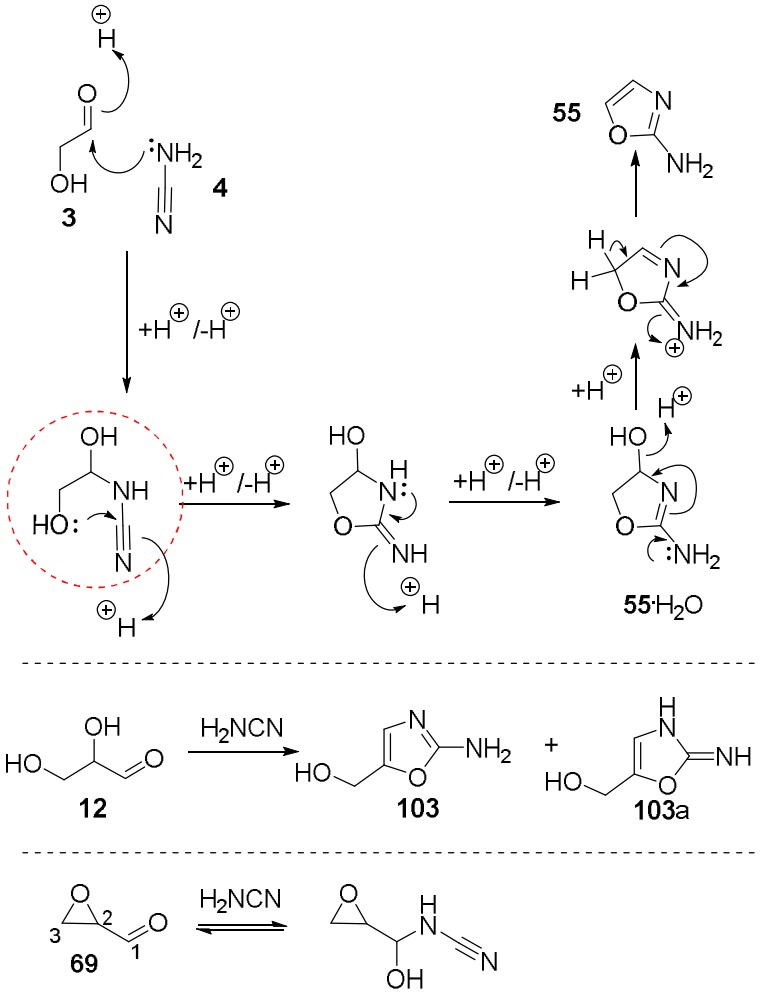


*Scheme 5.2: Scheme representing the outcomes of the three component reaction of cyanamide (**4**), glyceraldehyde (**12**) and glycolaldehyde (**3**) and subsequent possible combinations (desired reactions and products shown in blue, undesired products shown in red).*

This means that for aminooxazolines (**49**) to be selectively formed, yielding canonical nucleotides, a fractionation or temporally controlled addition of C_2 and

C_3 sugars needs to be imagined[289xiv], restricting geochemical environments in which desired chemistry can occur, rendering the desired outcome less probable. It is interesting to note that this selectivity problem might be solved by use of an aldehyde **12** analogue rather than direct use of **12**. The analogous nature of glycidaldehyde (**69**) with **12**, as well as the potential for **69** to react with 2-aminooxazole (**55**) to form aminooxazolines (**49**) selectively (*Section 4*), makes it an ideal candidate to replace **12** as a C_3 sugar synthon in the multi-component reaction to form **49**. Furthermore, we hypothesised that the C2 masked-hydroxyl group of **69** would be unavailable for cyclisation onto the nitrile of **4**, making it unavailable for oxazole formation, unlike **12** (*Scheme 5.3*). Lastly, **69** has the added benefit, compared to **12**, of being able to incorporate phosphate selectively at its C3 position, a potential that can be harnessed in the selective aqueous synthesis of 5′-phosphate aminooxazoline (**53**) from **69**, **55** and P_i, as demonstrated previously (*Section 4*).

[xiv] A problem currently investigated by the Powner lab by sequestration of aldehydes by 2-aminothiazole in the form of aminals[311], and the time-resolved nature of the reaction.



*Scheme 5.3: Reactions of glycolaldehyde (**3**), glyceraldehyde (**12**), and glycidaldehyde (**69**) with cyanamide (**4**) leading to oxazole formation. **3** and **12** lead to 2-aminooxazole (**55**; and its hydrate **55**·H_2O), and (2-aminooxazol-5-yl)methanol (**103**), respectively, while **69** does not react with **4** to give oxazoles, due to the masked-C2 hydroxyl. The key cyclisation step is shown in the red circle.*

At first we decided to study the formation of epoxide **69** from acrolein (**1**) in the presence of the aforementioned components of the reaction (**3** and **55**) in order to determine the limits of the feasibility of such a reaction. We had previously studied the epoxidation of **1** in water (*Section 2*) (140mM 1:1 **1**:H_2O_2, pH 8.5) which yielded **69** efficiently and selectively (>90%). We next sought to investigate the compatibility of this reaction with the presence of glycolaldehyde

(**3**). An equimolar reaction of **3**:H_2O_2 (140mM, in water) to which was added **1** (1 eq, pH 9) was prepared and monitored by ^{1}H NMR spectroscopy (*Figure 5.1*). After 6 h, the reaction showed near quantitative conversion of **1** to **69** (92%) in the presence of **3**.

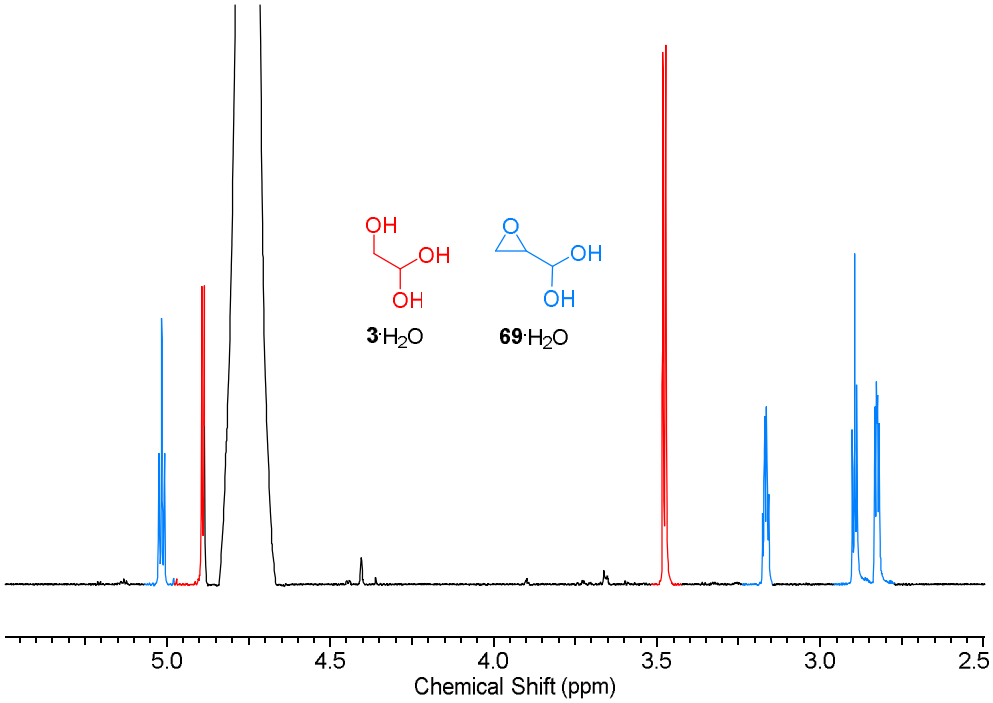


*Figure 5.1: ^{1}H NMR spectrum (600 MHz, H_2O/D_2O, 9:1, 5.50-2.50 ppm) of glycidaldehyde (**69**) production (blue) in the reaction of acrolein (**1**) and H_2O_2 in the presence of glycolaldehyde (**3**; red) (in water, pH 9, 140mM, 6 h).*

Similarly, encouraging results were obtained when acrolein (**1**) was added to a solution of 2-aminooxazole (**55**) and H_2O_2 (140mM, in water, pH 9), as all **1** was consumed within 15 min and 57% **69** was detectable (*Figure 5.2*). The lower yield can be accounted for by the fact that any **69** formed can immediately react with **55**[xv], which was reinforced by the observation that only 22% of **55** was left at this time point (based on ^{1}H NMR integration and compared to initial **55** concentration).

[xv] The reaction of **69** and **55** is detailed in *Section 4*.

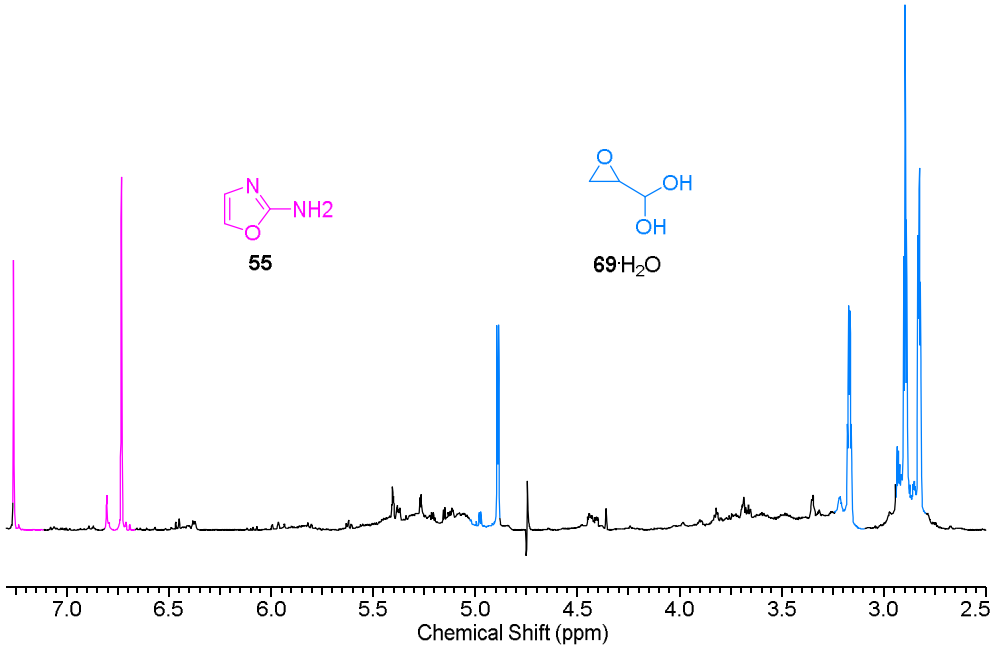
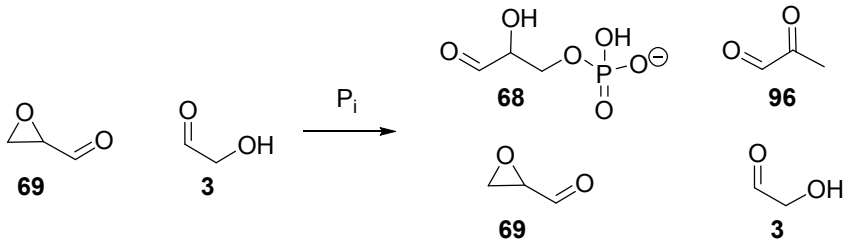


*Figure 5.2: ^1H NMR spectrum (600 MHz, H_2O/D_2O, 9:1, 7.30-2.50 ppm) of glycidaldehyde (**69**) production (blue) in the reaction of acrolein (**1**) and H_2O_2 in the presence of 2-aminooxazole (**55**; pink) (in water, pH 9, 140mM, 15 min).*

Having established that there is scope for *in-situ* synthesis of epoxide **69** from acrolein (**1**), it was then important to study the reactivity of **69** with the compounds of interest in this multicomponent reaction, namely glycolaldehyde (**3**) and cyanamide (**4**).

First we studied the reaction of glycidaldehyde (**69**; 128mM) and glycolaldehyde (**3**; 1.1 eq, 0.5M P_i buffer, pH 7, RT). This reaction showed only minimal reaction of **69** with some slow phosphorylation of **69** with P_i present occuring, yielding small amounts of glyceraldehyde-3-phosphate (**68**) over 4.5 days (15% **68** and subsequent E1cB elimination of phosphate leading to methylglyoxal (**96**), 5%; based on ^{1}H NMR integration). After 4.5 days at RT, **3** was unreacted (accounting for 50% of species in solution alongside 30% **69**, based on ^{1}H NMR integration, *Scheme 5.4*), leading us to conclude that **69** was stable in the presence of **3**.



*Scheme 5.4: Reaction of glycidaldehyde (**69**; 128mM) and glycolaldehyde (**3**; 140mM) in P_i buffer (0.5M) at pH 7 and RT, yielding a mixture of products over time, from nucleophilic P_i addition to **69** giving glyceraldehyde-3-phosphate (**68**), and its subsequent E1cB elimination giving methylglyoxal (**96**).*

A similar reaction of glycidaldehyde (**69**; 128mM) and cyanamide (**4**; 1.1 eq, 0.5M P_i buffer, pH 7) showed progressive degradation of **69** into methylglyoxal (**96**) and other unknown species over 9 days (*Figure 5.3*). The ^{1}H NMR analysis of this reaction showed presence of methylglyoxal **96**, which is most likely formed from the incorporation of P_i at the C3 position of **69** and subsequent elimination (as detailed in *Section 3*), accounting for some of the loss of material. Despite the detrimental chemistry of **69** in the presence of **4** (*Figure 5.3*), excitingly, the reaction did not produce a significant amount of oxazole species, suggesting that the masked hydroxyl at the C2 position of **69** prevented any ring closure leading to oxazole compounds (*Scheme 5.3*, *vide supra*), thereby confirming our hypothesis that **69** is unable to react with **4** in the same way as **12**.

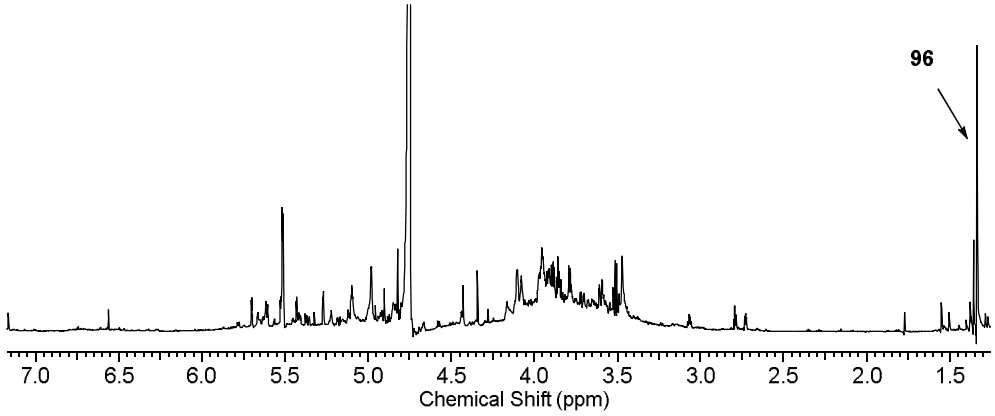


*Figure 5.3: ^{1}H NMR spectrum (600 MHz, H_2O/D_2O, 9:1, 7.15-1.25 ppm) of the incubation of cyanamide (**4**; 140mM) and glycidaldehyde (**69**; 128mM) in P_i buffer (0.5M, pH 7, RT) after 9 days, showing degradation of **69** to methylglyoxal (**96**) and other unknown species.*

We then investigated the competition of glyceraldehyde (**12**) and glycidaldehyde (**69**) for reaction with cyanamide (**4**), in the presence of P_i (based on the observation of P_i catalysed oxazole **55** formation from **3**, described above), in order to see which C_3 aldehyde preferentially reacted with **4** to form oxazoles. A reaction of **12**, **4** (140mM each) and **69** (128mM) in the presence of P_i (0.84M, pH 7, RT) showed formation of the oxazole species **103** (along with **103**a, *Scheme 5.3*). The total yield of the oxazole species **103** was 15% after 10 h, with 35% of starting material **12** left in solution (based on ^{1}H NMR integration with a standard, compared to initial **12** concentration). At this time point, a comparable amount of **69** was left in solution (39% based on ^{1}H NMR integration with a standard, compared to initial **69** concentration, *Figure 5.4, a.*). After a total of 3 days in the same conditions, there was an increase in **103** (and **103**a) oxazoles formed (21%) and no detectable **12** and **69** left (*Figure 5.4, b.*). Based on the incubation of **69** and **4** detailed previously, the observed depletion of **69** in this competition experiment suggested that **69** was reacting *via* a different unknown mechanism than **12** and was not stable under in the conditions tested.

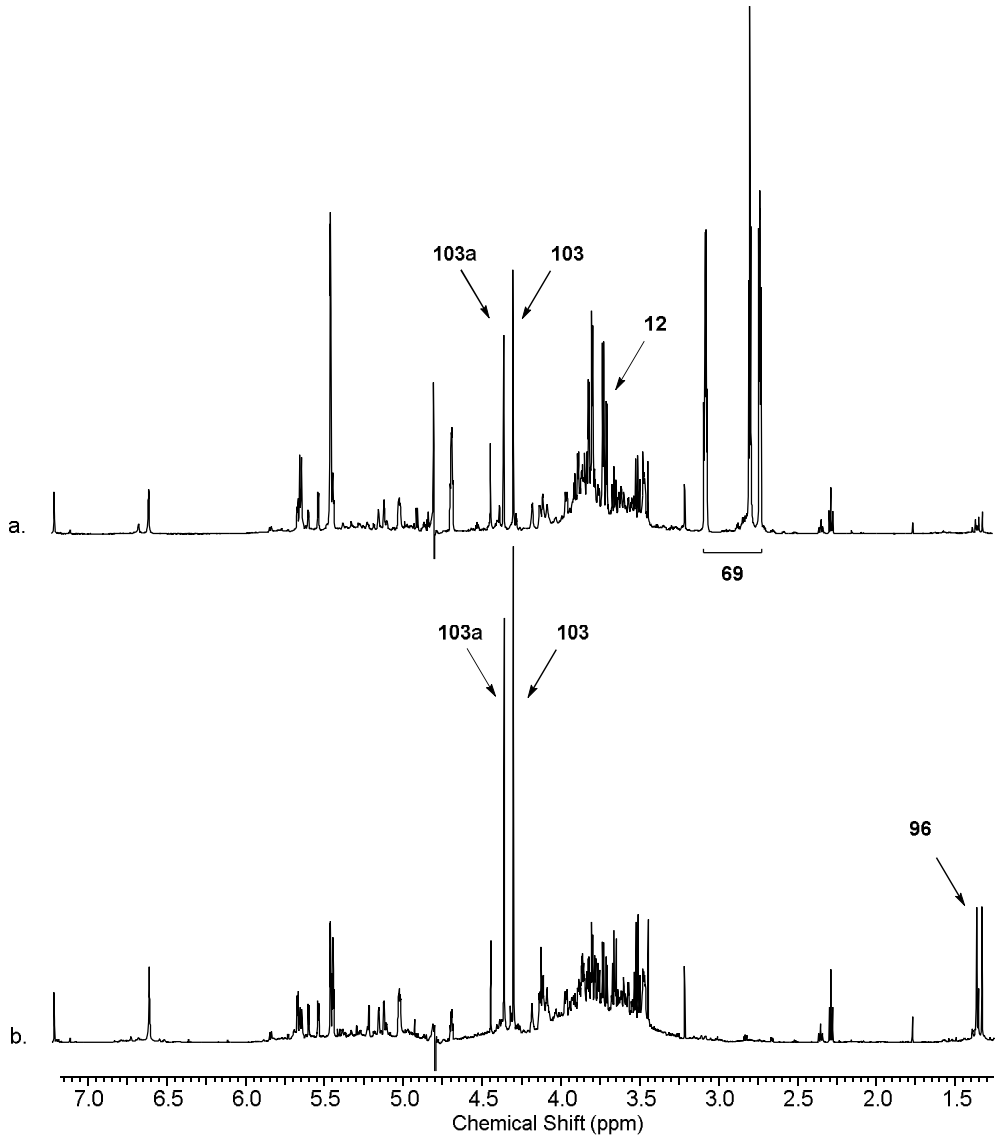


*Figure 5.4: ^{1}H NMR spectra (600 MHz, H_2O/D_2O, 9:1, 7.15-1.25 ppm) of the competition experiment of glyceraldehyde (**12**), cyanamide (**4**; 140mM each) and glycidaldehyde (**69**; 128mM) in P_i buffer (0.84M, pH 7, RT): a. after 10 h; b. after 3 days, showing degradation of **69** and formation of (2-aminooxazol-5-yl)methanol (**103**).*

The results obtained from the incubation of glycidaldehyde (**69**) with cyanamide (**4**) or glycolaldehyde (**3**) opened up the potential to use **69**, instead of glyceraldehyde (**12**), in a multicomponent reaction involving **3**, **4** and P_i, in the hopes that **3** and **4** would yield **55** selectively in the presence of **69**, after which the reaction of **69** and **55** with P_i could occur (as detailed in *Section 4*), yielding

pentose aminooxazolines 5′-phosphate (**53**) selectively, unobstructed by a mixture of by-products (like that depicted in *Scheme 5.2*). Therefore, it was important to verify that **55**, the oxazole formed from **4** with **3**, could form in the presence of epoxide **69**. In order to limit degradation of **69** through its phosphorylation to glyceraldehyde-3-phosphate (**68**) and subsequent phosphate elimination to methylglyoxal (**96**), we first studied the reaction of **69** (128mM), **3** (168mM) and **4** (168mM) without phosphate (in water, pH 7). We expected slow formation of **55**·H_2O in these conditions, due to the lack of general/acid base catalysis from P_i, leading to a sluggish cyclisation and aromatisation. Indeed, we observed a maximum of 40% **55**·H_2O being formed after 6 h at RT (based on ^{1}H NMR integration with a standard, based on initial **3** concentration, *Figure 5.5, b*.).

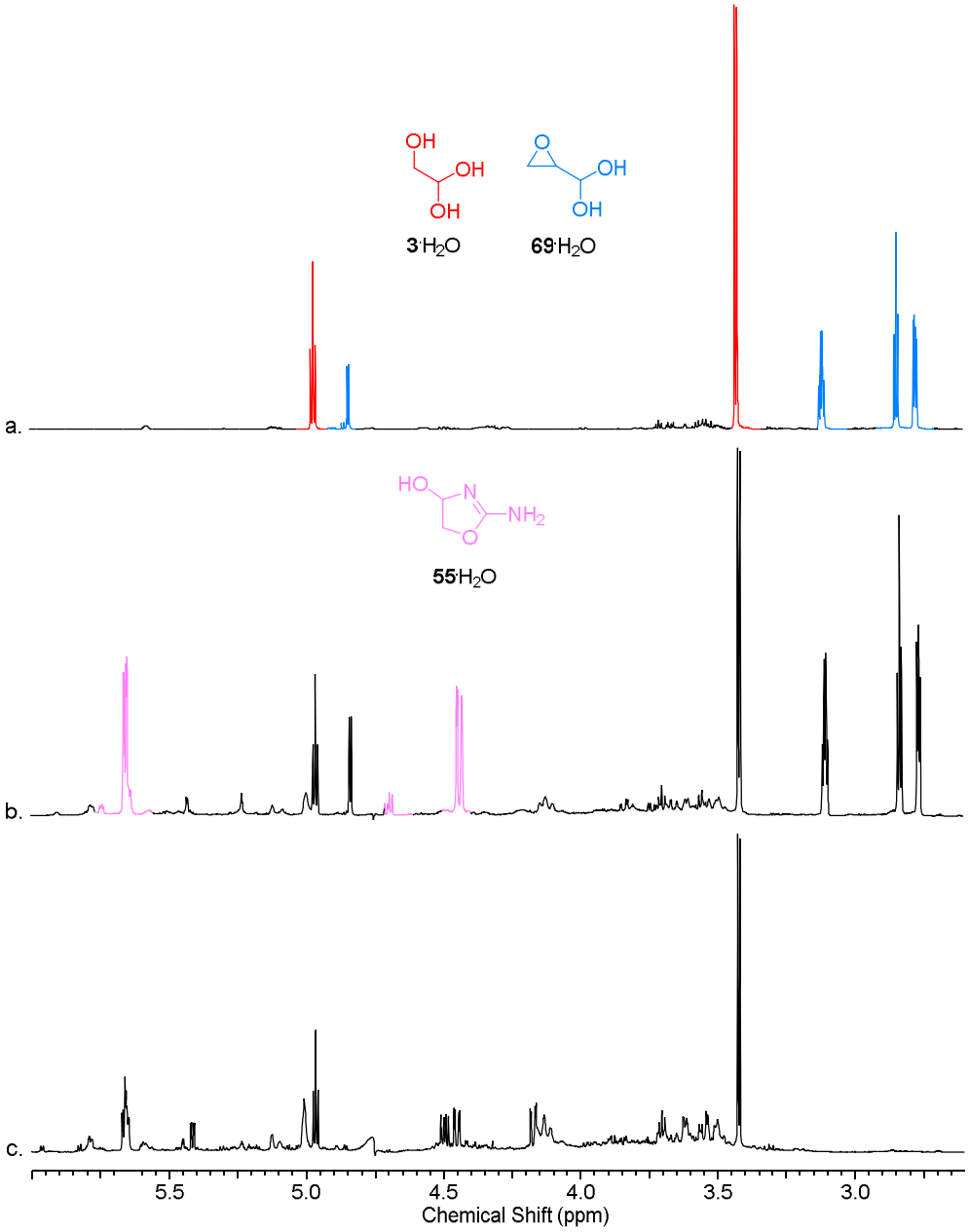


*Figure 5.5: ^{1}H NMR spectra (600 MHz, H_2O/D_2O, 9:1, 6.00-2.60 ppm) of the reaction of glycidaldehyde (**69**[xvi]; blue), glycolaldehyde (**3**; red) and cyanamide (**4**; in water, pH 7, 128:168:168mM respectively) showing 2-aminooxazole hydrate (**55**·H_2O; pink) production and **69** degradation after a. 0.3 h; b. 6 h and; c. 26 h.*

[xvi] In water, both glycidaldehyde (**69**) and glycolaldehyde (**3**) sit primarily (<95%) as hydrates in water and so are referred to by their aldehydic name throughout for simplification.

However, the formation quickly plateaus at this point (*Figure 5.6*), most likely due to loss of cyanamide (**4**) in its unknown detrimental reactions with epoxide **69**. Nevertheless, this showed potential for 2-aminooxazole (**55**) formation in the presence of **69**. Perhaps in a reaction containing P_i, the formation of **55** would be more efficient and reach completion. The question is: will **55** form before the stock of **69** is depleted? And will we observe selective formation of 5'-phosphorylated aminooxazolines **53** (our goal in this multicomponent reaction, *Scheme 5.5*)?

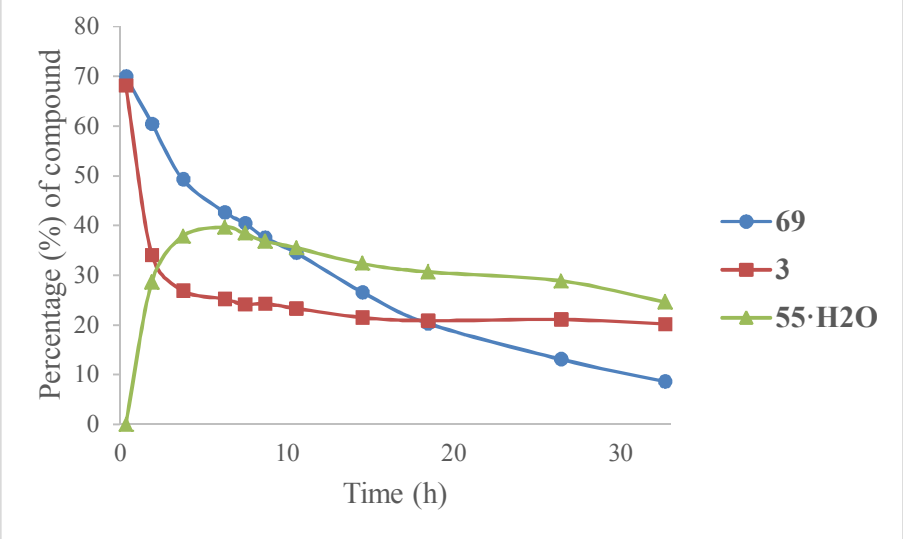
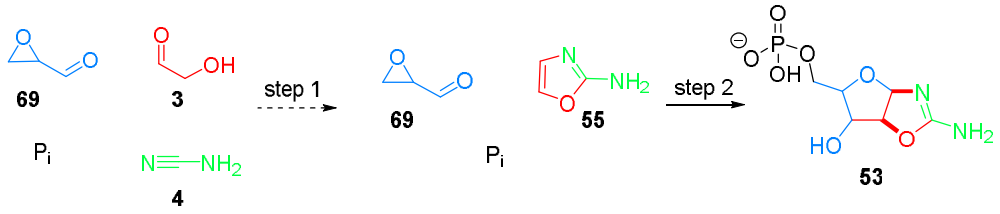


*Figure 5.6: 2-aminooxazole hydrate (**55**·H_2O) production and glycidaldehyde (**69**) degradation in the reaction of **69**, glycolaldehyde (**3**) and cyanamide (**4**; in water, pH 7, 128:168:168mM respectively). Percentages are calculated based on internal standard and the initial concentration of each compound in solution (**55**·H_2O is calculated based on starting concentration of **3**).*



*Scheme 5.5: Four-component reaction of glycidaldehyde (**69**), glycolaldehyde (**3**), cyanamide (**4**) and P_i. Step 1 (hypothetical): Initial reaction to form 2-aminooxazole (**55**) selectively in the presence of **69** and P_i. Step 2 (observed): Three-component reaction of **69**, **55**, and P_i leading to pentose aminooxazolines 5'-phosphates (**53**).*

In order to explore the multicomponent reaction of glycidaldehyde (**69**), glycolaldehyde (**3**), cyanamide (**4**) and P_i, the reaction was analysed with a consistent concentration of epoxide **69** (128mM) and various ratios of **3**, **4** and P_i, at RT[xvii] and at a range of pH (5-9). Lower pH reactions didn't show much improvement compared to the reaction described above without P_i as limited 2-aminooxazole (**55**) synthesis was observed. Indeed, **55** formation is only observed around neutral pH or higher (in the presence of P_i) due to the need for deprotonation of **55**·H_2O to yield **55**[33]. This essential step therefore can't be achieved at lower pH as less P_i dianion (an efficient general acid catalyst) is present. At neutral pH reactions the formation of **55** was observed, along with the formation of aminooxazoline species, as indicated by the characteristic doublets in the 6.0-5.7 ppm region of ^{1}H NMR (*Figure 5.7*, marked with *). ^{31}P-^{1}H HMBC analysis of the sample showed clear correlation between the H-C5′ region of 5'-phosphate aminooxazolines (**53**; 4.0-3.5 ppm) and a phosphorus signal (7.5-5.5 ppm; *Figure 5.8*), showing clear evidence of phosphorylation at the 5′-position. This signal is comparable to the ^{31}P-^{1}H HMBC correlation seen in the 3-component reaction of **69**, **55** and P_i described in *Section 4* (*Figure 4.10*, *a.* p. 134; *Scheme 5.5*, step 2). These results were very encouraging, as they suggest there is **53** formation, our desired outcome of this 4-component reaction (*Scheme 5.5*).

[xvii] Some reactions were carried out at 40°C but showed more detrimental **69** degradation and therefore were not pursued further.

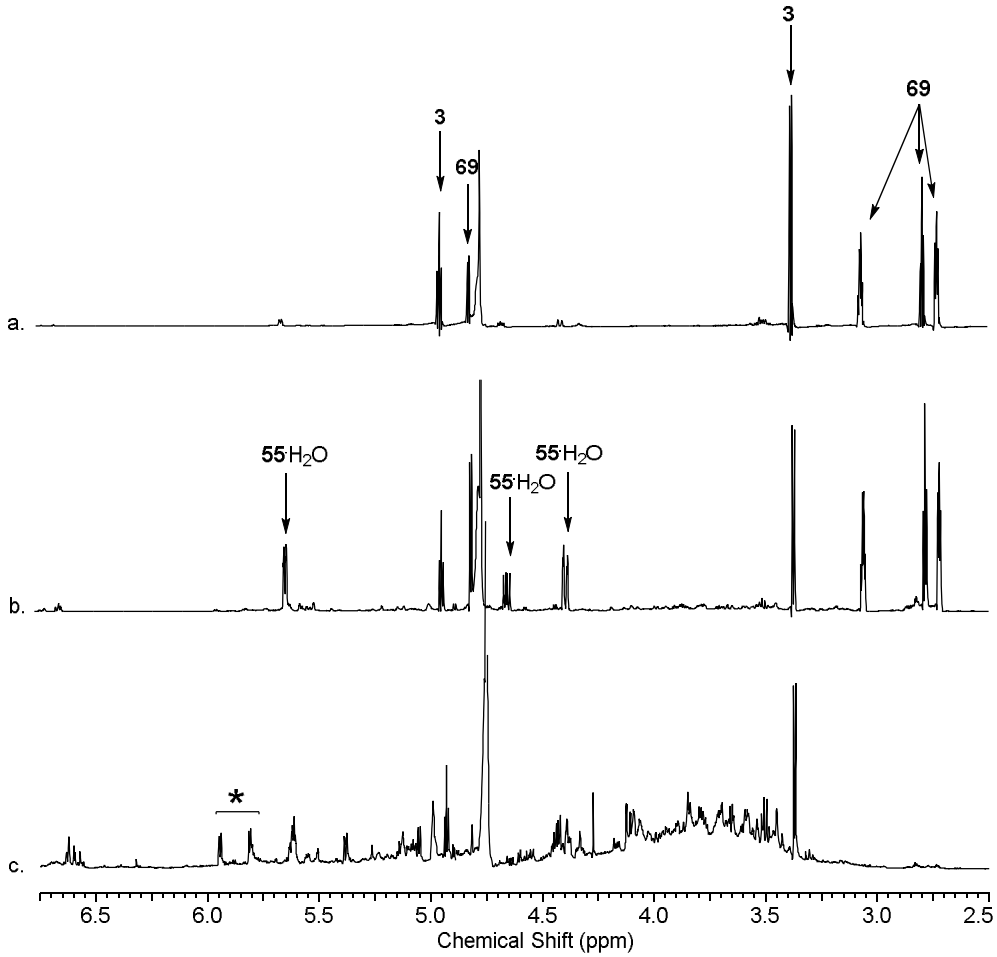


*Figure 5.7: ^{1}H NMR spectra (600 MHz, H_2O/D_2O, 9:1, 6.75-2.50 ppm) showing the formation of 2-aminooxazole hydrate (**55**·H_2O), the consumption of glycidaldehyde (**69**) and the formation of aminooxazolines (annotated with *) in the reaction of glycolaldehyde (**3**), cyanamide (**4**) and glycidaldehyde (**69**; 1:1:1 140mM) in P_i buffer (0.5M) at pH 7 after a: 0.3 h; b. 10 h; c. 226 h.*

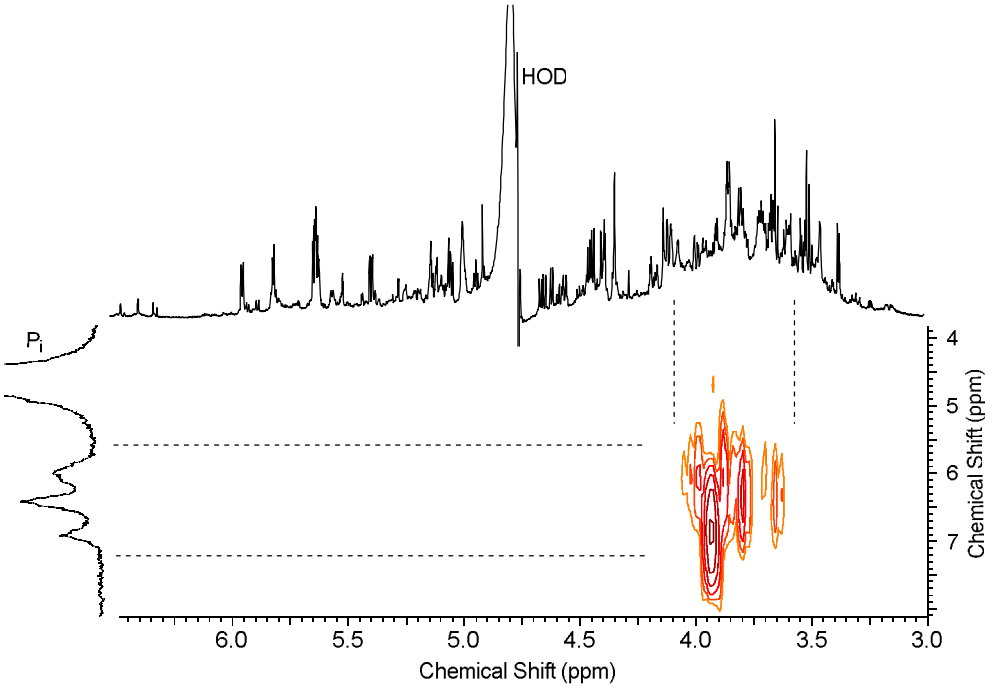


*Figure 5.8: ^{1}H-^{31}P HMBC (300 MHz, H_2O/D_2O 9:1) showing coupling between phosphorus and protons in the 3.5-4 ppm region, evidence of successful phosphorylation at the 5′ position of aminooxazolines, in the reaction of glycolaldehyde (**3**; 140mM), cyanamide (**4**; 168mM), and glycidaldehyde (**69**; 128mM) in P_i buffer (0.5M) at pH 7 after 7 days.*

Taking into account the complexity that can arise from a 4-component reaction, it was important to devise a spiking strategy that would take into account all of the main possible outcomes of the reaction, namely sequential spiking with pentose aminooxazolines (**49**), their 5′-phosphate form **53** and tetrose aminooxazolines (**104**) (we also looked at the possible formation of 3′-hydroxymethyl tetrose aminooxazolines (**105**) which can be obtained through hydrolysis of glycidaldehyde (**69**) to glyceraldehyde (**12**), isomerisation to dihydroxyacetone (**8**) and subsequent reaction with 2-aminooxazole (**55**)). Spikings (using synthetic samples) were first carried out on the multicomponent reaction involving **12** (**12**:**3**:**4**: P_i in water, pH 7, 140:168:168:500mM respectively), in order to have a clear comparison of the chemistries at play, and their selectivity or lack thereof (*Figure 5.9*). These spikings showed that the main species formed in these conditions were, as expected (*Scheme 5.2*): **104**, **49** and oxazoles **103**.

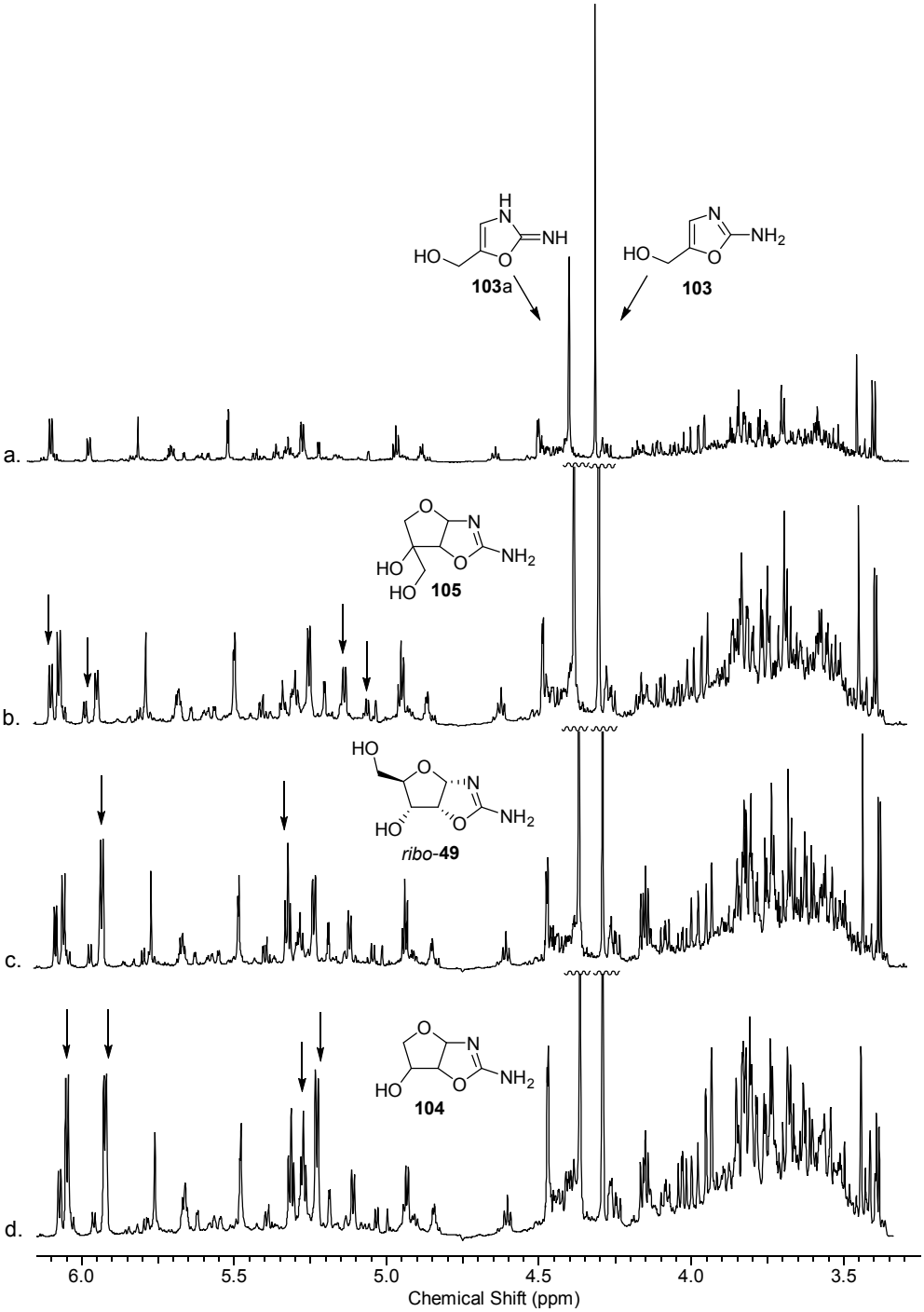
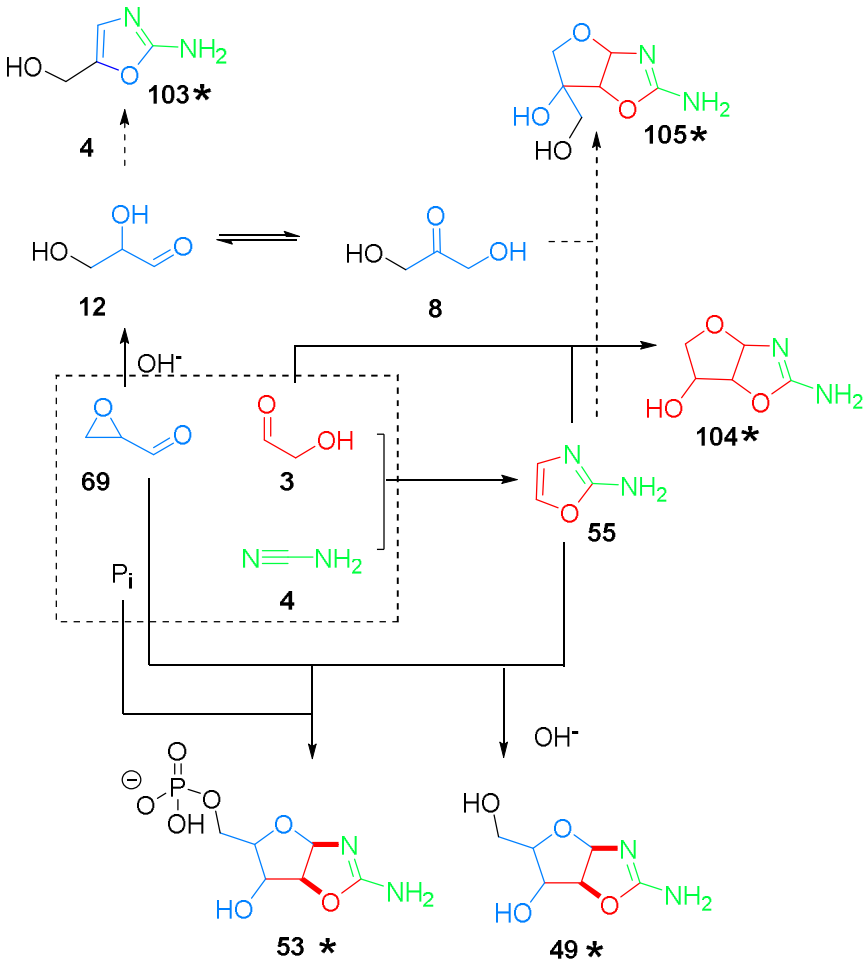


*Figure 5.9: ^{1}H NMR spectra (600 MHz, H_2O/D_2O, 9:1, 6.15-3.25 ppm) a. Multicomponent reaction of glyceraldehyde (**12**; 140mM), glycolaldehyde (**3**; 168mM), cyanamide (**4**; 168mM) and P_i (500mM) in water, pH 7. Lyophilised after 6 days, resuspended in D_2O and pH to 5. NMR acquired after 24 h. Spiked with authentic samples of: b. 3'-hydroxymethyl tetrose aminooxazolines (**105**); c. riboaminooxazoline (ribo-**49**); d. tetrose aminooxazolines (**104**).*

Next we analysed a similar multicomponent reaction using epoxide **69** as a C_3 aldehyde instead of glyceraldehyde (**12**; pH 7, **69**:**3**:**4**:P_i 128:168:168:500mM) at RT, once it had gone to completion (after 5 days, *Figure 5.10, a.*). Spiking this reaction with the same compounds as above (*Scheme 5.6*) showed that the main species in solution were tetrose aminooxazolines (**104**) and pentose aminooxazolines (**49**) (*Figure 5.10, b.* and *Figure 5.11, c.*). However, knowledge that ^{31}P-^{1}H HMBC revealed the presence of correlation between the H-C5′ region of aminooxazolines to a phosphate group (*Figure 5.8*; *vide supra*) in this series of experiments gave us confidence that some 5'-phosphate aminooxazolines (**53**) were being formed. This reaction showed a slight improvement in selectivity, thanks to the lack of chemistry through **12**: no observed formation of oxazoles **103** (hypothetically formed from the hydrolysis of **69** to **12** and subsequent reaction with **4**) and no observed branched oxazolines **105** (hypothetically formed from the hydrolysis of **69** to **12**, isomerisation of **12** to its ketose form **8** and subsequent reaction with **55**; *Scheme 5.6*). In spite of this slight improvement in selectivity, and the added benefit of phosphorylation to **53** (compared to **12** reactions) this 4-component reaction was still lacking in selectivity for our wanted product **53**. This was most likely due to the degradation of **69** in the presence of **4** (*vide supra*), leading to an excess of **3** in solution and ultimately leading to an increase in tetrose aminooxazolines (**104**) formation (*Scheme 5.6*).



*Scheme 5.6: Four-component reaction of glycidaldehyde (**69**), glycolaldehyde (**3**), cyanamide (**4**) and P_i (within the dashed box) and the potential outcomes of the reaction (solid lines: observed; dashed lines: hypothetical). Spiked compounds represented with *.*

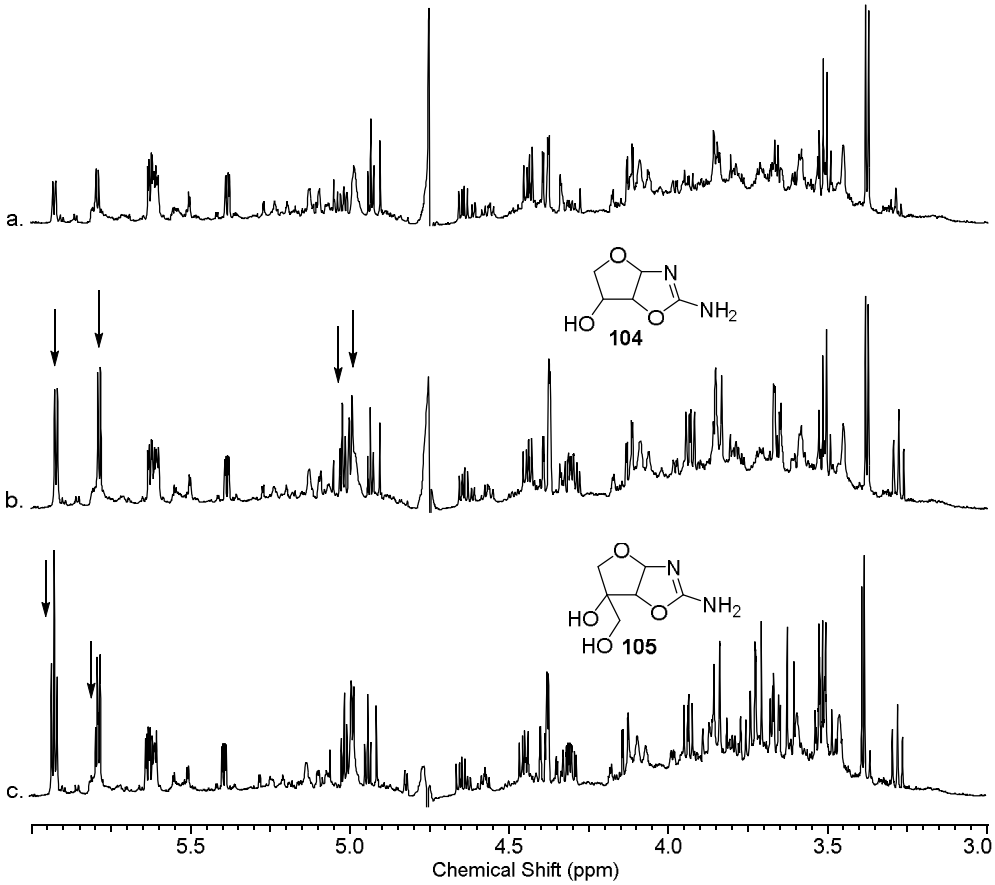


*Figure 5.10: ^{1}H NMR spectra (600 MHz, H_2O/D_2O, 9:1, 6.00-3.00 ppm) a. Multicomponent reaction of glycidaldehyde (**69**; 128mM), glycolaldehyde (**3**; 168mM), cyanamide (**4**; 168mM) and P_i (500mM) in water, pH 7. NMR acquired after 5 days. Spiked with authentic samples of: b. tetrose aminooxazolines (**104**); c. 3'-hydroxymethyl tetrose aminooxazolines (**105**).*

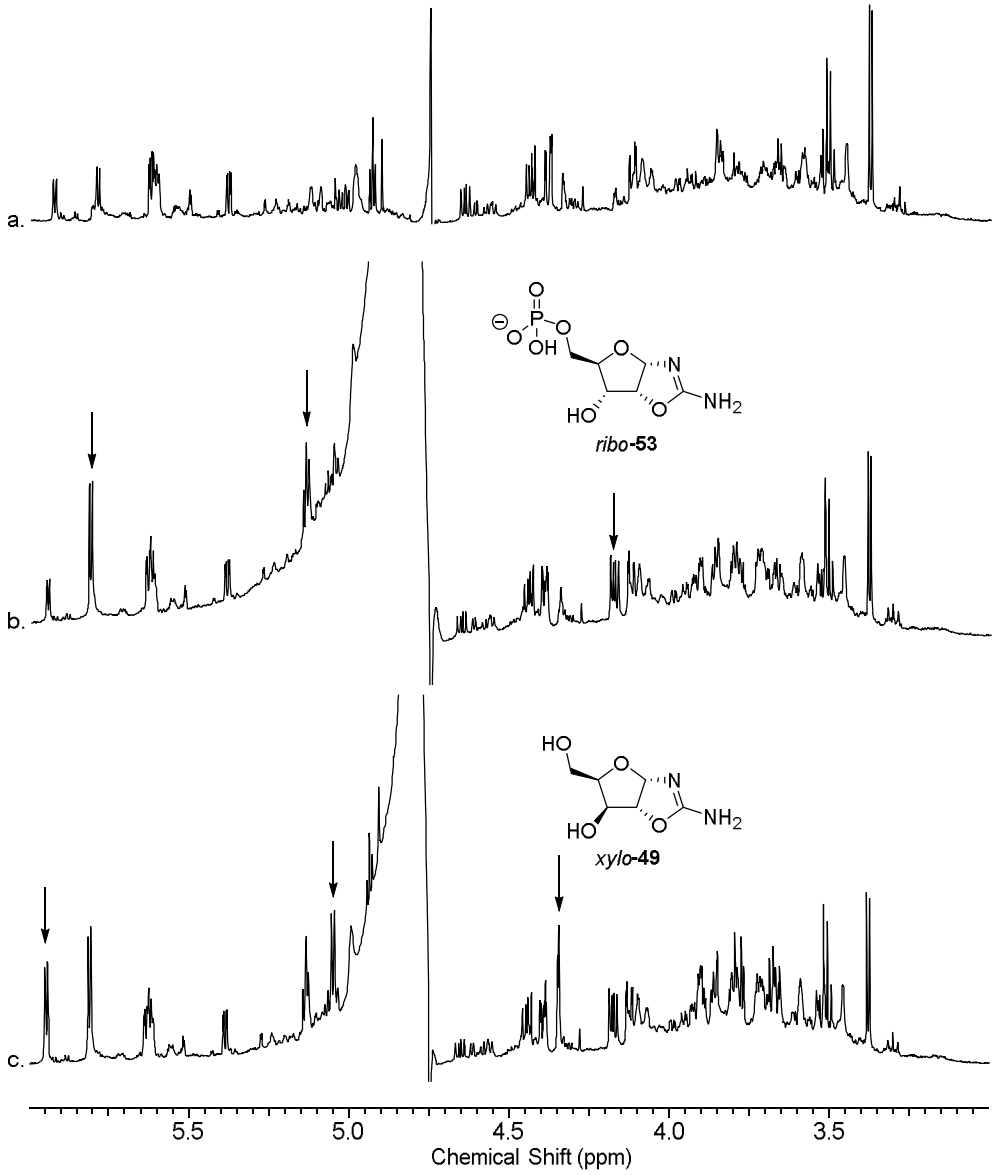


*Figure 5.11: ^1H NMR spectra (600 MHz, H_2O/D_2O, 9:1, 6.00-3.00 ppm) a. Multicomponent reaction of glycidaldehyde (**69**; 128mM), glycolaldehyde (**3**; 168mM), cyanamide (**4**; 168mM) and P_i (500mM) in water, pH 7. NMR acquired after 5 days. Spiked with authentic samples of: b. riboaminooxazoline 5'-phosphate (ribo-**53**); c. xyloaminooxazoline (xylo-**49**).*

The promising lack of new oxazoles **103** led us to investigate various conditions that might led to an improved selectivity for our phosphorylation pathway. Repeating the reaction with an excess of **69** with respect to **3** (**69**:**3** 2:1) showed an improvement with regards to amounts of tetrose aminooxazolines (**104**) formed, as they were not obviously present in reaction (*Figure 5.12*, *a.* lack of a triplet at

3.5-3.4 ppm and *b.* inconclusive spike). Again ^{31}P-^{1}H HMBC revealed the presence 5'-phosphate aminooxazolines (**53**), however spikings with authentic samples of *ribo*-**53** and *xylo*-**53** were inconclusive (*Figure 5.13*), leading us to speculate that the formation of **53** was only minor and accompanied with other aminooxazoline species, notably non-phosphorylated **49**. Difficulties in interpretation lie in the overlap of signals along with the sheer number of compounds that can be formed from this reaction (4 diastereomers of **53**, 4 diastereomers of **49**, 2 diastereomers of **104**). This experimental difficulty along with the lack of an obvious selectivity for **53** led us to conclude that while there is potential for some C_3 selectivity and phosphorylation to **53** through the use of our epoxide **69**, the benefits gained are not enough to completely solve the aforementioned C_2 and C_3 selectivity problem.

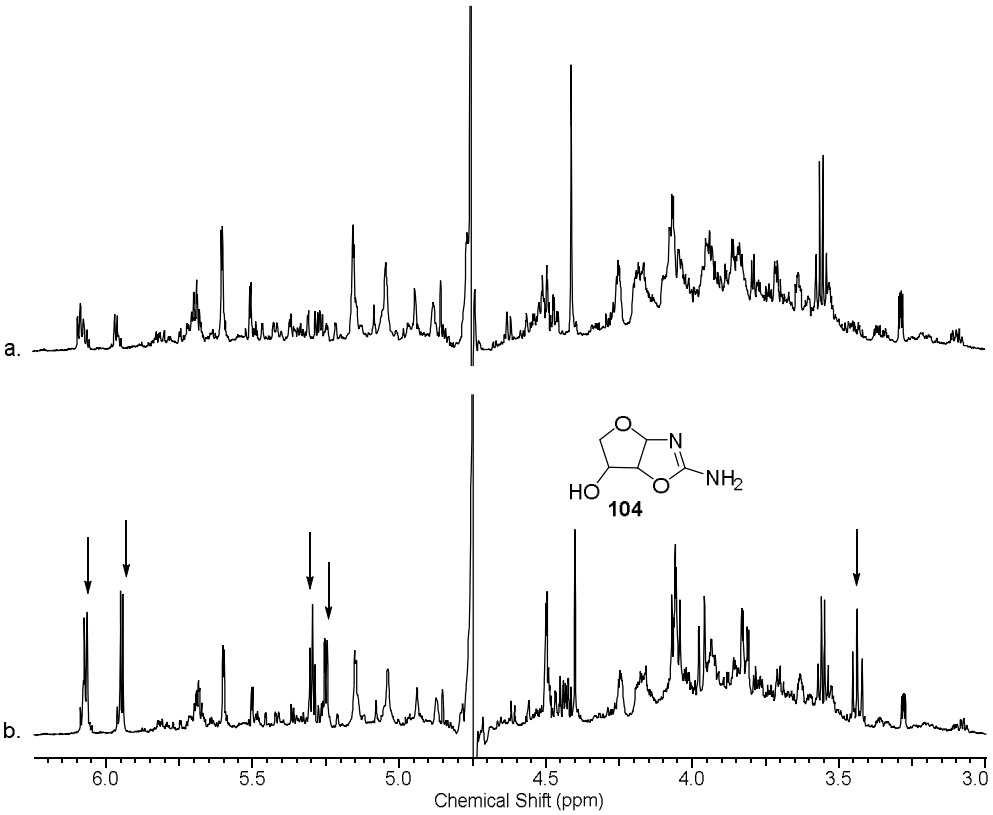


*Figure 5.12: ^{1}H NMR spectra (600 MHz, H_2O/D_2O, 9:1, 6.25-3.00 ppm) a. Multicomponent reaction of glycidaldehyde (**69**; 86mM), glycolaldehyde (**3**; 43mM), cyanamide (**4**; 86mM) and P_i (430mM) in water, pH 7, 7 days. NMR acquired after reaction was adjusted to pH 5 for 2 h. b. spiked with an authentic sample of tetrose aminooxazolines (**104**).*

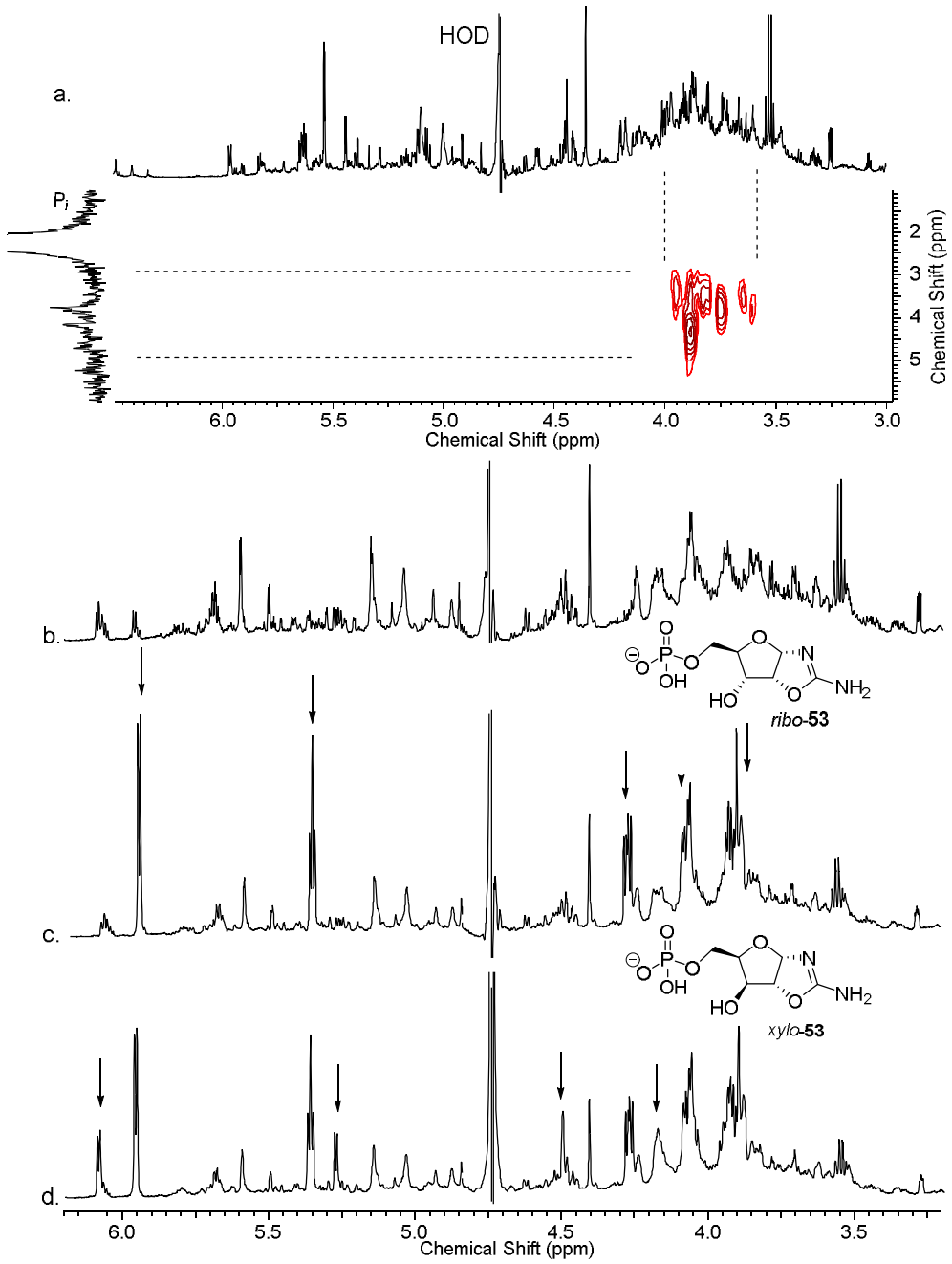


Figure 5.13: a. ^{1}H-^{31}P HMBC (400 MHz, H_2O/D_2O 9:1) showing coupling between phosphorus and protons in the 3.5-4 ppm region, evidence of successful phosphorylation at the 5′ position of aminooxazolines, in the reaction of *glycidaldehyde (**69**; 86mM), glycolaldehyde (**3**; 43mM), cyanamide (**4**; 86mM) and P_i (430mM) in water, pH 7, after 7 days. b. ^{1}H NMR (600 MHz, H_2O/D_2O, 9:1, 6.25-3.00 ppm) spectra acquired after the reaction was adjusted to pH 5 for 2 h. Spiked with authentic samples: c. riboaminooxazoline 5′-phosphate (ribo-**53**); d. xyloaminooxazoline 5′-phosphate (xylo-**53**).*

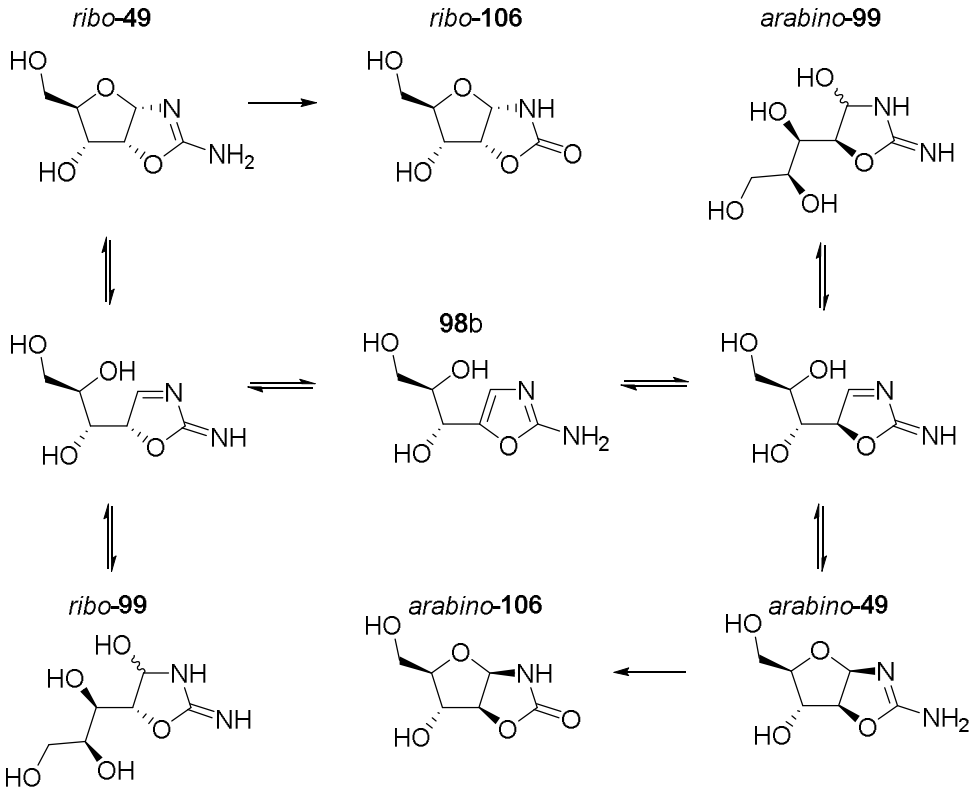
In conclusion, these results show that while there is potential for epoxidation of acrolein (**1**) to glycidaldehyde (**69**) in the presence of essential components to the prebiotic synthesis of 5′-phosphate aminooxazolines (**53**), (such as glycolaldehyde (**3**) and 2-aminooxazole (**55**)) the difficulty in obtaining the needed sequential reactions for selective **53** synthesis (without outside control from the chemist) hinders the prospect of a 4 component synthesis of **69**, **3**, cyanamide (**4**), and P_i. Therefore, we propose that epoxide **69** should be regarded as a useful compound in prebiotic synthesis due to its potential for nucleophilic insertion at the C3 position, notably of phosphate (*Section 3*), in water under mild conditions, thereby leading to selective formation of 5′-phosphorylated aminooxazolines (**53**) in the 3 component reaction of **69**, **55** and P_i (*Section 4*). Unfortunately, its inherent reactivity as an epoxide, especially with **4**, limits its use in this 4 component reaction and does not help solve the selectivity problem observed in the indiscriminate reaction of **12** and **3** with **4**, as had been hoped. We suppose one might find an elegant way of by-passing the limitations observed in our reactions, however no further work was done on this matter from our part.

6. From aminooxazolines 5′-phosphates to nucleotides

6.1. Anomerisation studies of aminooxazolines 5′-phosphates

Following their work on the prebiotic synthesis of activated pyrimidine ribonucleotides[33], the Sutherland group investigated the behaviour of one of the intermediates *en route* to pyrimidines, pentose aminooxazolines (**49**), in the medium used in the synthesis: phosphate buffer[309]. They found that phosphate catalysed the interconversion of *arabino*-**49** and *ribo*-**49**. The main reason for the study was to find an explanation for the observed depletion of one of the diastereomers over time, *lyxo*-**49**. This depletion helped change the proportion of the diastereomers in solution, effectively enriching the solution in their desired diastereomer *arabino*-**49**. *Arabino*-**49** is the diastereomer needed for the correct stereochemistry of the end product β-ribocytidine-2',3'-cyclic phosphate (β-**43**; and β-ribouridine-2',3'-cyclic phosphate (β-**57**); a point to be returned to later). The results of this incubation study showed that when *ribo*-**49** (86mM) is incubated in 1M P_i buffer for 6 days at neutral pH and at RT, no major changes are observed[309]. However, increasing the temperature to 60°C (at pH 7) leads to the formation of new species over the same time frame, specifically the aromatic open-chain species **98**b (4%), *arabino*-**49** (3%), oxazolidinones *arabino*-**106** (21%) and *ribo*-**106** (65%) and starting material *ribo*-**49** (7%; *Scheme 6.1*)[309]. Interestingly, the same incubation in deuterated P_i buffer (pD 7) at 60°C saw a collapse of the H-(C1′) signal for *ribo*-**49** from a doublet to a singlet. This led to a suggested mechanistic pathway (*Scheme 6.1*), which they then confirmed by synthetic preparation of one of the pathway intermediates **98**b followed by its incubation, leading to the same mixture of products being formed, with a preference for *ribo* compounds *ribo*-**49** and *ribo*-**106**. This preference for *ribo*-**49** over *arabino*-**49** has been attributed to the *endo*-cyclic orbital overlap ($n_O \rightarrow \delta^*_{C\text{-}N}$ interaction) possible in this α-anomer, leading to stabilisation through a

pseudo-anomeric effect. This hypothesis is in part confirmed by crystal X-ray diffraction data, which is consistent with the observed stability of *ribo*-**49**[308].



*Scheme 6.1: Pathway of the anomerisation and hydrolysis of riboaminooxazoline (ribo-**49**) and arabinoaminooxazoline (arabino-**49**), as observed by Powner et al.*[309] *in their anomerisation studies in 1M P_i buffer.*

We hypothesised our 5′-phosphate aminooxazolines (**53**; obtained from the multicomponent reaction of glycidaldehyde (**69**), 2-aminooxaole (**55**) and P_i; described in *Section 4*) would behave in a similar manner to **49** with regards to anomerisation, and that we could take advantage of the increased stability of *ribo*-**49** over *arabino*-**49** to simplify our mixture of four diastereomers (45% *ribo*-**53**, 16% *arabino*-**53**, and 39% *xylo*-**53**; *Section 4*, *Table 4.2*). Furthermore, we invistigated the anomerisations of **53** in P_i buffer (as described above) and in water, as we proposed that the **53** phosphate could effictively replace the P_i due to its general acid/base properties and high effetive molarity. Pure samples of *ribo*-**53**, *arabino*-**53** and *xylo*-**53** were prepared by a synthetic route (by incubation

of the corresponding 5-phosphate sugars (**52**) with cyanamide (**4**) in the presence of ammonia), and each was incubated (**53**: 0.1M) in water or P_i buffer (1M) at a range of temperature (RT to 60°C) in order to assess their behaviour. The C2′ stereochemistry of the *ribo*- and *xylo*-derivatives both proved to be stable in the presence and absence of P_i buffer at RT. The incubation of *ribo*-**53** at RT (*Figure 6.1*) showed that 74% of *ribo*-**53** remained after 7 days in 1M P_i buffer, pH 7 (based on ^{1}H NMR integration relative to other species in solution, *Table 6.1,* entry 2) and only a minor amount of aromatisation to **100**b occured (20%; singlet at 6.63 ppm). No anomerisation to *arabino*-**53** was observed at RT.

Entry n°	Temperature (°C)	Conc. of P_i (M)	^{1}H NMR integration relative to other species integrated (%)					
			ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
1	RT	-	82.5	0.8	1	0.72	12.9	2.7
2	RT	1	73.9	1.3	2.2	0.3	19.5	2.8
3	40	-	40	6.6	7.3	1.5	39.5	5.1
4	40	1	38.1	4.6	20.7	3.2	30	3.4

*Table 6.1: Composition of species following the incubation of riboaminooxazoline-5'-phosphate (ribo-**53**) at pH 7, after 7 days, at RT or 40°C, with and without the presence of phosphate buffer.*

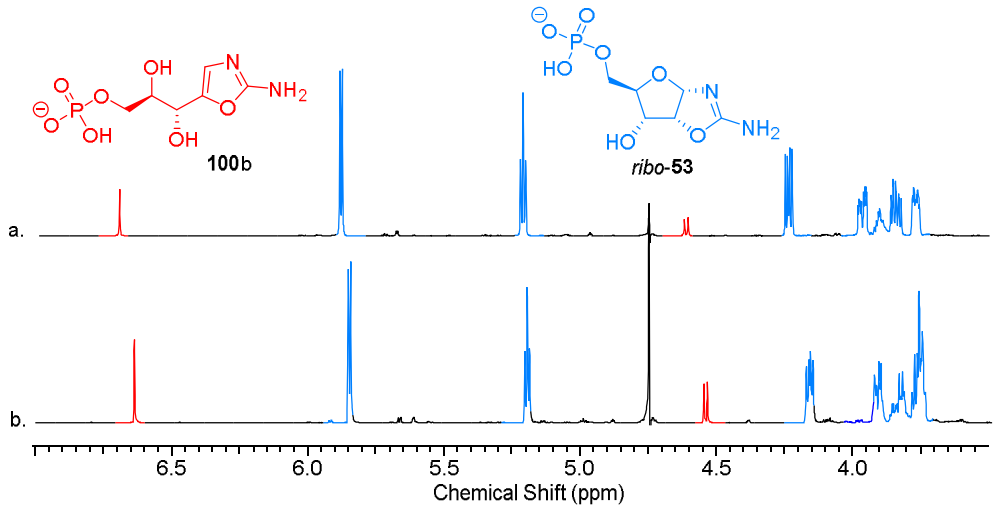


*Figure 6.1: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 1:1, 7.00-3.50 ppm) of the reaction of riboaminooxazoline-5'-phosphate (ribo-**53**; blue) at pH 7, RT; a. after 7 days in H_2O/D_2O 1:1; b. after 7 days in 1M P_i buffer. Showing no obvious change in composition of species, with only minor amounts of aromatisation to open-chain species **100**b (red) taking place. Characteristic peaks for the compounds have been highlighted.*

The incubation of *xylo*-**53** at RT (*Figure 6.2*) showed similar results with 83% of *xylo*-**53** remaining after 7 days in 1M P_i buffer, pH 7 (based on ^{1}H NMR integration relative to other species in solution, *Table 6.2,* entry 2) and again only a minor amount of aromatisation to **100**a (12%; singlet at 6.63 ppm) was observed. No anomerisation to *lyxo*-**53** was observed.

Entry n°	Temperature (°C)	Conc. of P_i (M)	^{1}H NMR integration relative to other species integrated (%)					
			xylo-**53**	*lyxo*-**53**	*xylo*-**107**	*lyxo*-**107**	**100**a	**101**a
1	RT	-	77.9	0	0.4	0	20	1.7
2	RT	1	83.3	0	2.1	0	12.3	2.3
3	40	-	56.8	0	9.2	0	29.2	4.8
4	40	1	44.1	0	23.7	0	28.5	3.7

*Table 6.2: Composition of species following the incubation of xyloaminooxazoline-5'-phosphate (xylo-**53**) at pH 7, after 7 days, at RT or 40°C, with and without the presence of phosphate buffer.*

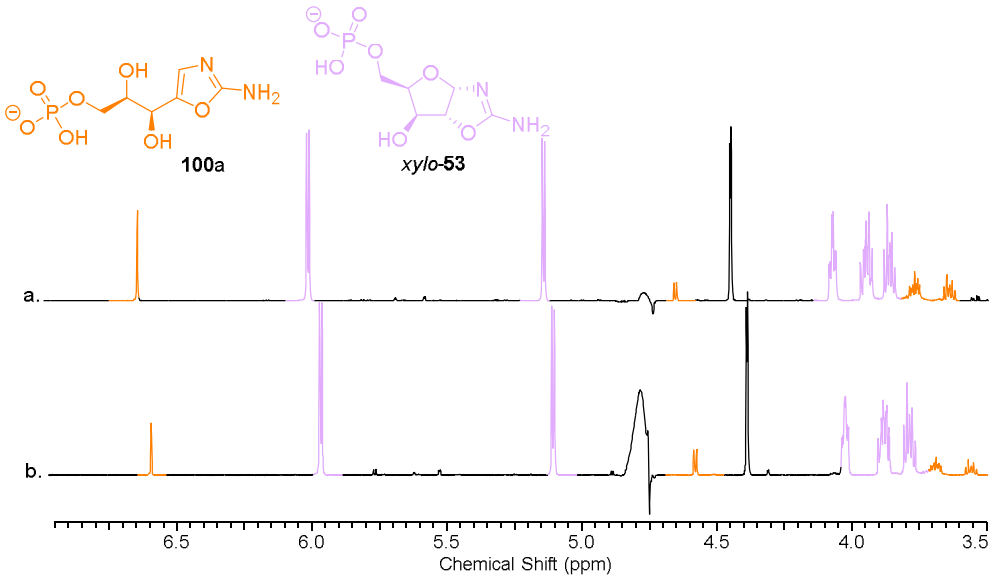
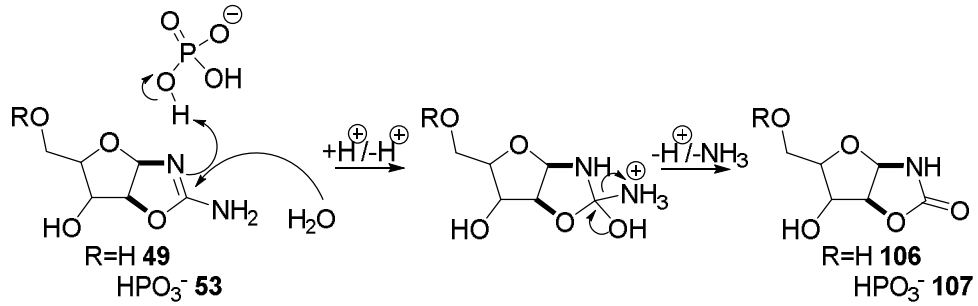


*Figure 6.2: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 1:1, 7.00-3.50 ppm) of the reaction of xyloaminooxazoline-5'-phosphate (xylo-**53**; pink) at pH 7, RT; a. after 7 days in H_2O/D_2O 1:1; b. after 7 days in 1M P_i buffer. Showing no obvious change in composition of species, with only minor amounts of aromatisation to open-chain species **100**a (orange) taking place. Characteristic peaks for the compounds have been highlighted.*

When *ribo*-**53** and *xylo*-**53** were incubated at 40°C for 7 days (in the absence of P_i, pH 7), there was a marked increased in the aromatic species formed (40% **100**b, 29% **100**a respectively, by ^{1}H NMR, relative to other species in solution). However, incubation of *ribo*-**53** did not lead to significant formation of *arabino*-**53** (1.5% *arabino*-**53**, and no *lyxo*-**53** was detected in the incubation of *xylo*-**53**). However, a portion of the aminooxazolines *ribo*-**53** and *xylo*-**53** degraded to their respective oxazolidinones 5′-phosphate (**107**) (7% *ribo*-**107**, 9% *xylo*-**107** respectively, *Table 6.1/6.2,* entry 3). In P_i buffer (1M, pH 7), the proportion of oxazolidinone formation increased significantly, due to the general acid/base action of phosphate (now in a 11:1 excess; *Scheme 6.2*) leading to hydrolysis of the C2-NH_2 of **53**[309] (21% *ribo*-**107**, 24% *xylo*-**107** respectively, after 7 days, *Figure 6.3,* confirmed by spiking with authentic samples, *Figure 6.4* and *6.5*). Additionally, significant amounts of aromatisation were also observed (30% **100**b, 29% **100**a respectively, *Table 6.1/6.2,* entry 4).



Scheme 6.2: Mechanism for the formation of oxazolidinones from aminooxazolines by general acid/base action of P_i.

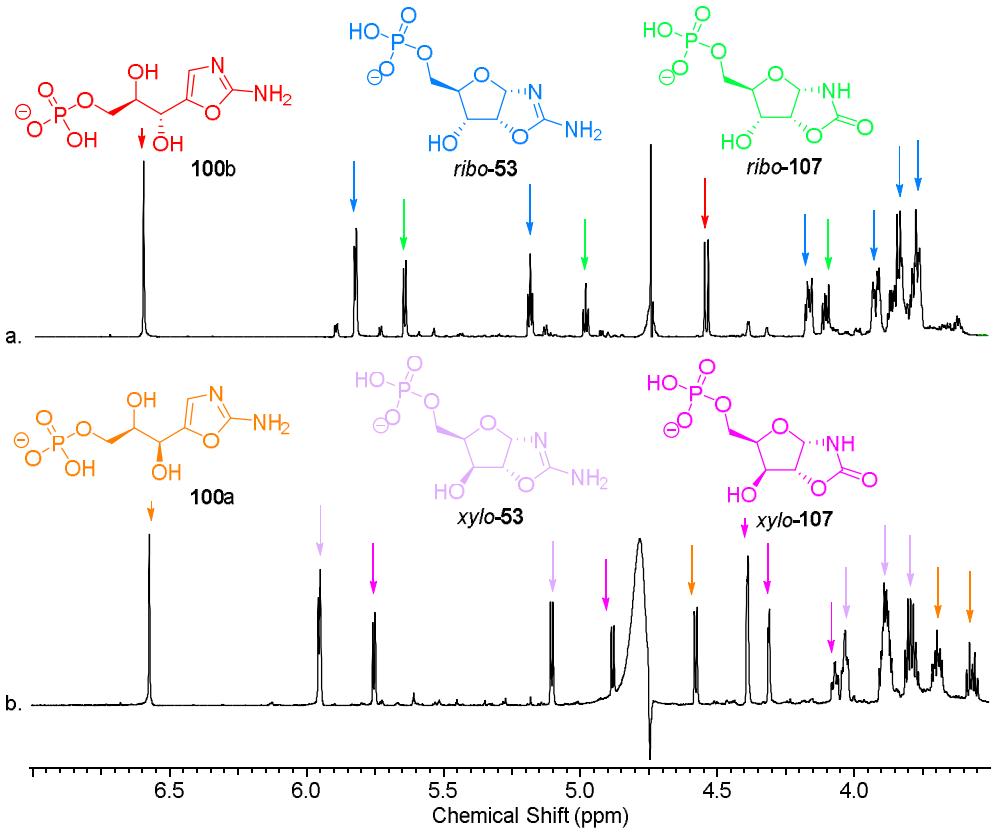


*Figure 6.3: ^1H NMR spectra (600 MHz, H_2O/D_2O 1:1, 7.00-3.50 ppm) of the reaction of aminooxazolines 5'-phosphate (**53**) at pH 7, 40°C, 1M P_i buffer after 7 days. a. ribo-**53**; and b. xylo-**53**. Showing formation of their corresponding oxazolidinones 5'-phosphate (**107**): ribo-**107** and xylo-**107**.*

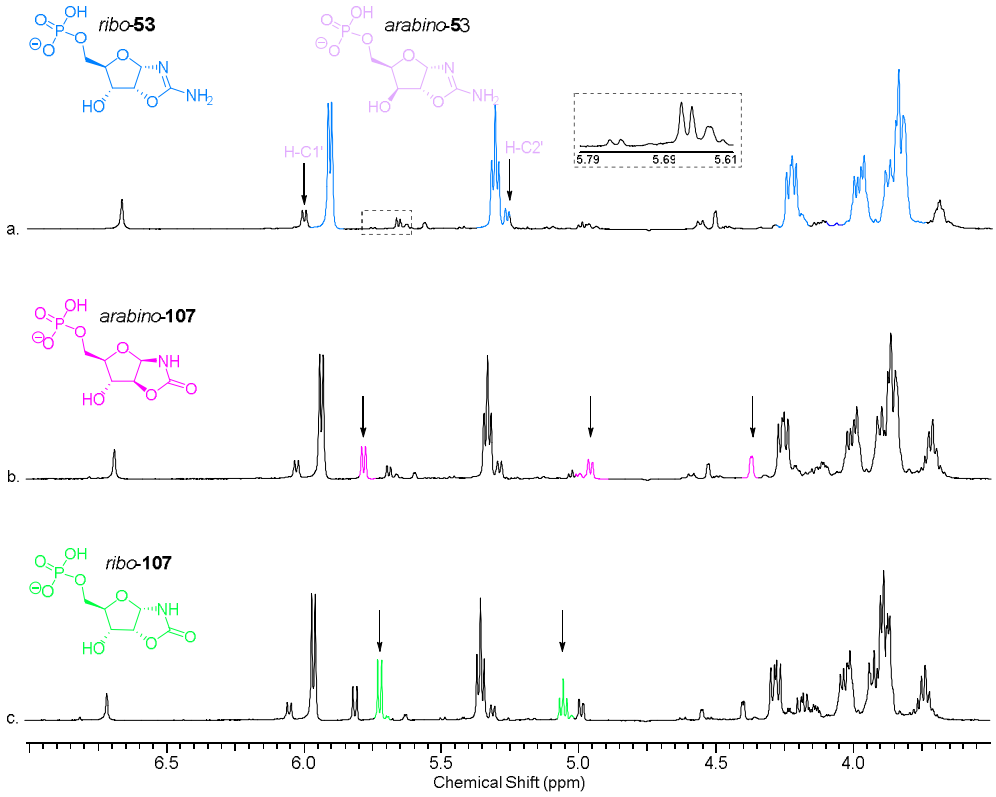


*Figure 6.4: ^{1}H NMR spectra (400 MHz, D_2O, 7.00-3.50 ppm) a. Incubation of riboaminooxazoline-5'-phosphate (ribo-**53**) at pH 6 at 40°C for 5 days in H_2O, followed by lyophilisation and subsequent suspension in D_2O; b. spiked with arabinooxazolidinone-5'-phosphate (arabino-**107**); c. spiked with ribooxazolidinone-5'-phosphate (ribo-**107**). Characteristic peaks for the compounds have been highlighted.*

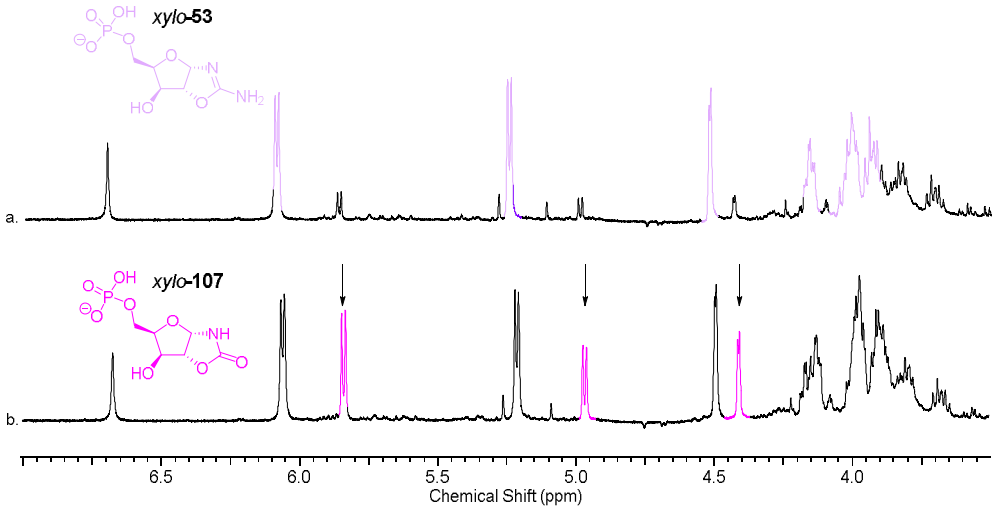
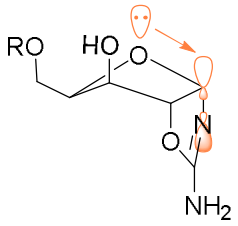


*Figure 6.5: ^{1}H NMR spectra (400 MHz, D_2O, 7.00-3.50 ppm) a. Incubation of xyloaminooxazoline 5′-phosphate (xylo-**53**) at pH 7 at 40°C for 7 days in H_2O, followed by lyophilisation and subsequent suspension in D_2O; b. spiked with xylooxazolidinone 5′-phosphate (xylo-**107**). Characteristic peaks for the compounds have been highlighted.*

Interestingly, only minor amounts of anomerisation were observed in these *ribo*-**53** experiments, with a maximum conversion to *arabino*- stereochemistry of 8% after 7 days at 40°C (*arabino*-**53** and *arabino*-**107** combined; 1M P_i, pH 7, *Table 6.1,* entry 4). This corroborates the observation that *ribo*-**49** was found to be stable with only minor conversion to *arabino*-**49** (in P_i buffer) as reported by Powner *et al.*[309] (*vide supra*). No discernible anomerisation from *xylo*-**53** to the *lyxo*-stereochemistry was observed. The *xylo*-isomer, with the same α-constitution as the *ribo*-isomer, can also benefit from the increased stability afforded to it by the *pseudo*-anomeric effect found from the axial configuration of the C1′ (*Figure 6.6*), and as so we expected it to act much like its *ribo*- counter part. The increased preference for *xylo*-**53** over *lyxo*-**53** when compared to *ribo*-**53** and *arabino*-**53** will be discussed later. The results we obtained confirmed the hypothesis that the α-constitution governs the stability of the aminooxazoline, by showing that *xylo*-**53** is stable in the range of conditions tested.



R= PO_3^{2-} *xylo*-**53**

Figure 6.6: Proposed orbital interaction $n_{O4} \rightarrow \delta^*_{C1'\text{-}N1}$ *leading to increased stability of xyloaminooxazoline 5′-phosphate (xylo-**53**).*

In order to confirm the mechanisms at play, *ribo*-**53** was incubated in D_2O (pD 7, 40°C), where the collapse of the H-(C1′) doublet to a singlet was observed over 7 days. This suggested that *ribo*-**53** was opening, and the H-C2′ was removed, giving access to aromatic **100**b (as seen with the presence of the 6.6 ppm singlet, *Figure 6.7*) which is planar at C2′, followed by deuteration at the C2′ and subsequent ring-closure yielding the most stable form, *ribo*-**53** (*Figure 6.7*). These observations are in-line with the observations made by Powner *et al.*[309] and confirmed that the mechanisms are analogous (*Scheme 6.3*).

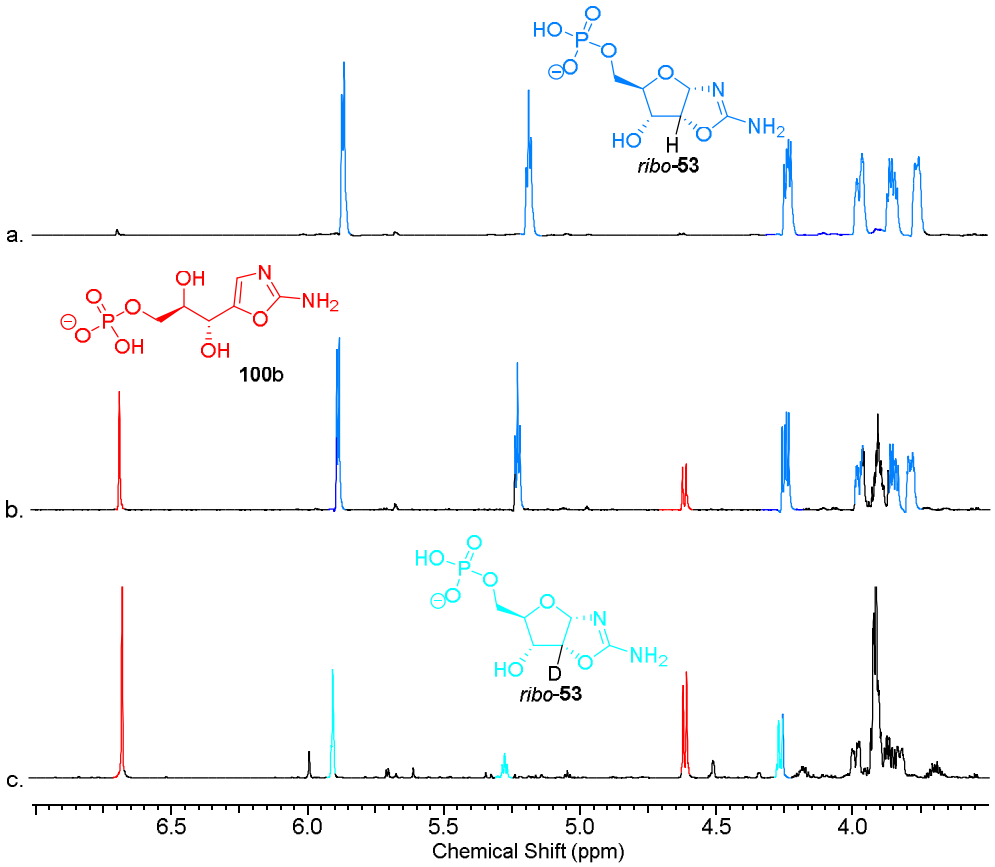
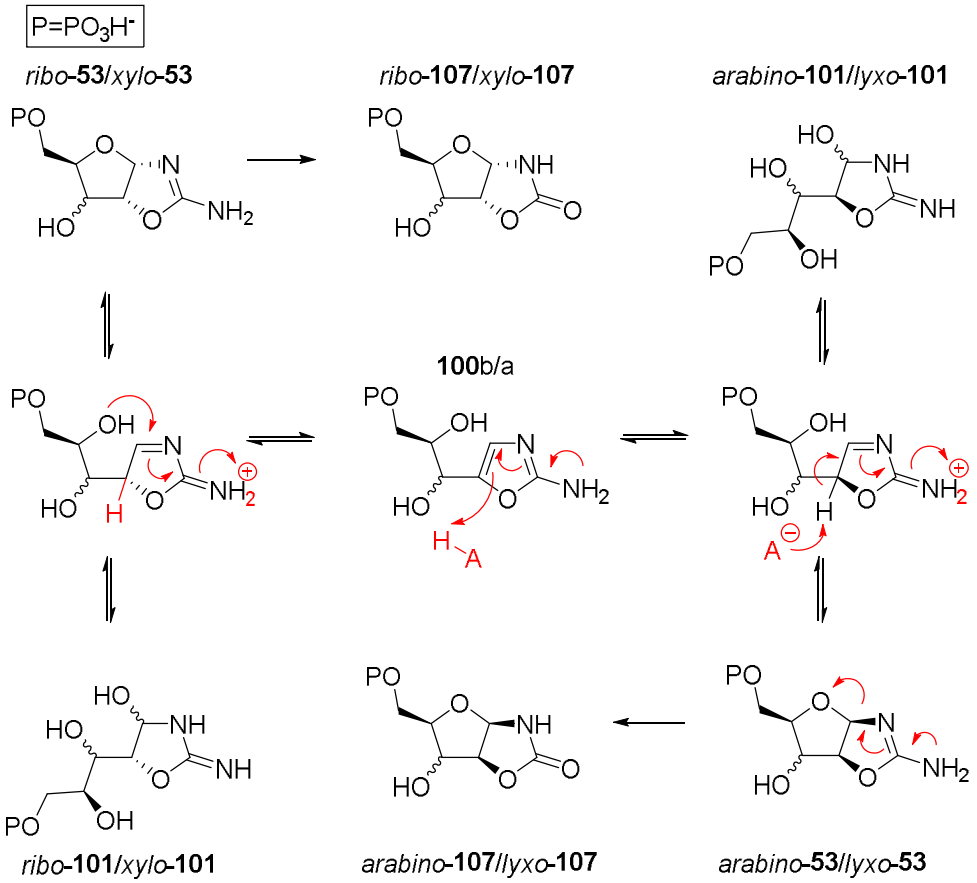


*Figure 6.7: ^{1}H NMR spectra (600 MHz, D_2O, 7.00-3.50 ppm) of the incubation of riboaminooxazoline-5'-phosphate (ribo-**53**) at 40°C, pH 7 showing the H/D exchange at the C2' position over time: a. 30 min; b. 1 day; c. 7 days. Characteristic peaks for the compounds have been highlighted.*



*Scheme 6.3: Pathway of the anomerisation and hydrolysis of ribo-/xylo- aminooxazoline 5'-phosphate ribo-**53**/xylo-**53** and arabino-**53**/lyxo-**53**, as observed in our anomerisation studies in H_2O or 1M P_i buffer, at a range of pH and temperatures. Mechanistic details for the preferred pathway are shown in red.*

As in the non-phosphorylated *ribo-*/*arabino*-**49** experiments carried out by Powner *et al*[309]., we expected that we would observe more chemistry in the case of the two β-anomers: *arabino*-**53** and *lyxo*-**53**. After synthetic preparation of *arabino*-**53**, we set out to analyse its behaviour at a range of pH, temperature and in the presence and absence of P_i. A marked increased in chemistry compared to the *ribo*-**53** incubations (*vide supra*) was already observed in the 'milder' set of conditions: absence of P_i, at RT and neutral pH. The small amount of anomerisation from *arabino*-**53** to *ribo*-**53** in these mild conditions (10% alongside 26% aromatisation to **100**b, after 7 days, *Table 6.3,* entry 1; *Figure 6.8, a.*) was already beyond the maximum obtained in the incubation of *ribo*-**53** (*vide*

supra; *Table 6.1*, entry 4 at 40°C, pH 7 in 1M P_i buffer). Significant conversion to *ribo*-**53** (20% relative to other species in solution, after 7 days, *Table 6.3*, entry 2) was observed in 1M P_i buffer, pH 7, still at RT (*Figure 6.8, b.*).

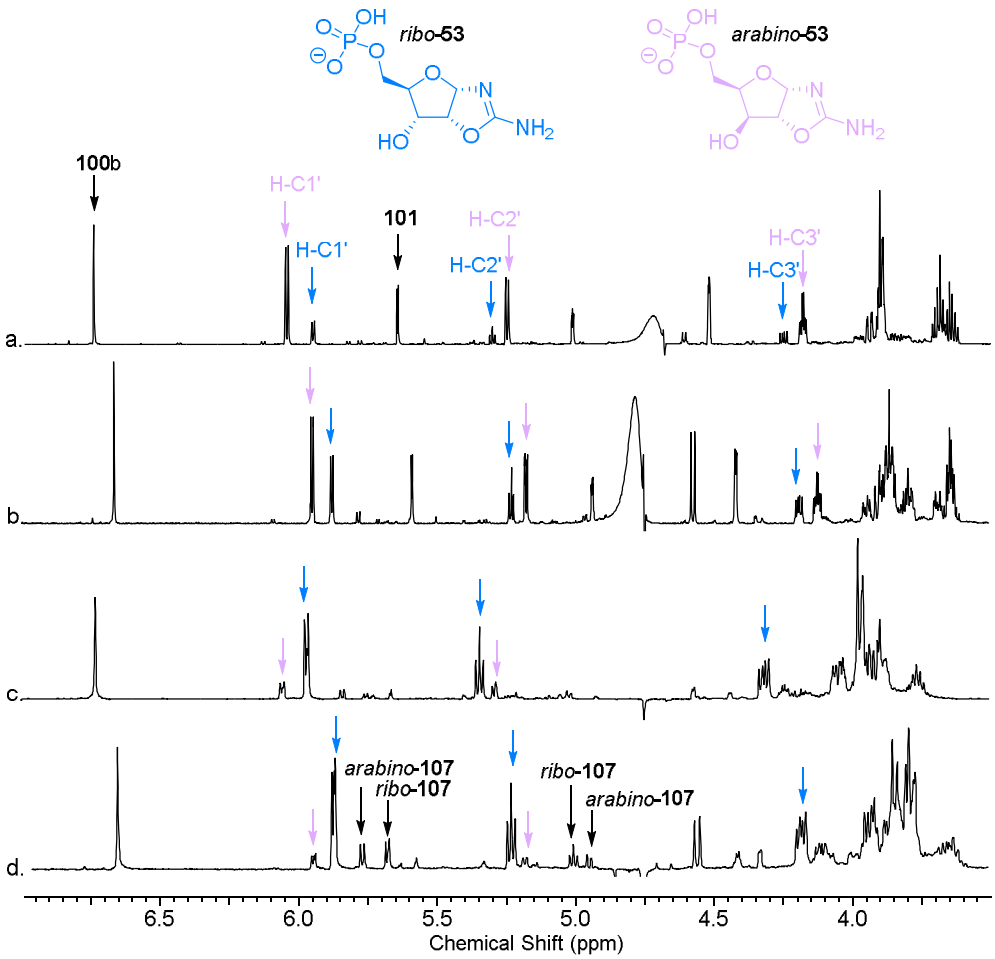


*Figure 6.8: ^{1}H NMR spectra (600 MHz, D_2O, 7.00-3.50 ppm) of the incubation of arabinoaminooxazoline-5'-phosphate (arabino-**53**) at pH 7 after 7 days: a. at RT, in H_2O; c. at RT, in 1M P_i buffer; c. at 40°C, in H_2O; d. at 40°C, in 1M P_i buffer. Showing evidence of anomerisation to ribo-**53**. The discernible peaks for arabino-**53** and ribo-**53** are pointed out.*

Entry n°	Temperature (°C)	Conc. of P_i (M)	^{1}H NMR integration relative to other species integrated (%)					
			arabino-**53**	*ribo*-**53**	*arabino*-**107**	*ribo*-**107**	**100**b	**101**b
1	RT	-	42.7	10.4	1	1.2	26.3	18.5
2	RT	1	33	19.7	3	1	24.9	18.4
3	40	-	10.8	45.1	5	3.4	32	3.7
4	40	1	6.9	44	9.1	11.3	25.6	3.3

*Table 6.3: Composition of species following the incubation of arabinoaminooxazoline-5'-phosphate (arabino-**53**) at pH 7, after 7 days, at RT or 40°C, with and without the presence of phosphate buffer.*

Incubation of *arabino*-**53** at 40°C led to a significant increase in the proportion of anomerisation from *arabino*-**53** to *ribo*-**53** of up to 45% after 7 days (*Table 6.3*, entry 3): extensive conversion was observed, even without the presence of P_i (*Figure 6.8, c.*). The main difference observed between experiments with and without P_i present, at 40°C, was the proportion of other species in solution. Specifically, the P_i buffer led to an increase in the relative amount of oxazolidinones (3% *ribo*-**107** and 5% *arabino*-**107** in water *vs* 11% *ribo*-**107** and 9% *arabino*-**107** in 1M P_i buffer) and a slight decrease in the proportion of aromatic species **100**b (32% in water to 26% in 1M P_i buffer; *Table 6.3*, entry 4; *Figure 6.8, d.*). After 14 days at 40°C, 1M P_i buffer, pH 7, *arabino*-**53** became a minor component of the reaction mixture (6%, little change compared to 7 days), while *ribo*-compounds accounted for 62% of the mixture (37% *ribo*-**53**, 25% *ribo*-**107**; *Figure 6.9*, bottom). Looking at the reaction over time (*Figure 6.9*) showed an increase in the singlet at 6.6 ppm for **100**b, and its subsequent decrease, suggesting that the aromatic species **100**b was the species converting to *ribo*-**53**, as depicted in the mechanistic pathway (*Scheme 6.3*, *vide supra*). Without P_i present, a sizable amount of the material sat in the aromatic form **100**b after 14 days (27% in water *vs* 17% in 1M P_i buffer) (*Figure 6.9*, top). The excess of phosphate in the P_i buffer (11:1) seemed to provide additional general acid/base catalysis for C2′ protonation leading to increased de-aromatisation from **100** to **53**, as well as increased hydrolysis of **53** to **107**.

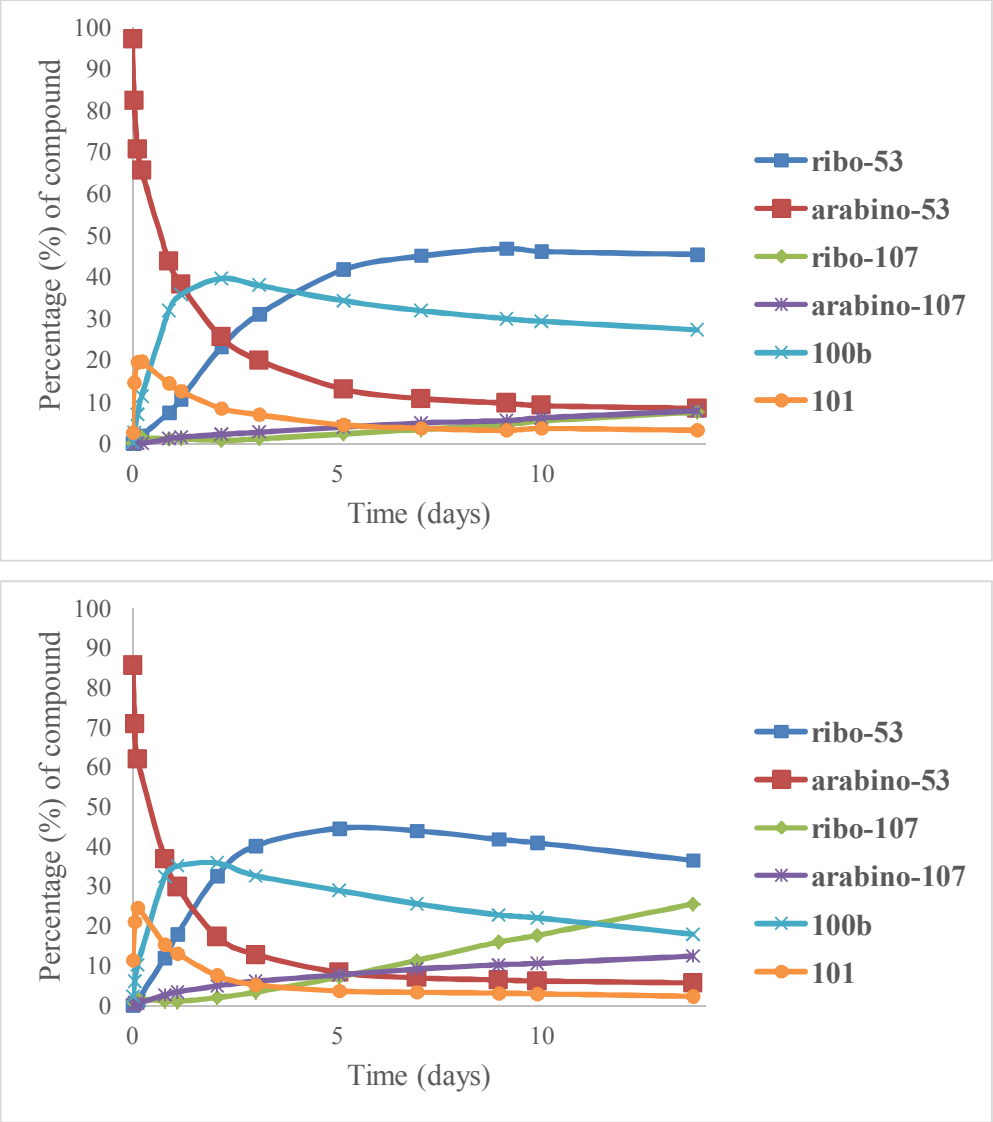


*Figure 6.9: Incubation of arabinoaminooxazoline-5'-phosphate (arabino-**53**; 0.1M) at pH 7, 40°C over time. Top: In H_2O/D_2O (1:1). Bottom: In 1M P_i buffer. Analysed by ^1H NMR spectroscopy. Percentages are reported as relative to identifiable species in solution.*

These results show a pronounced increase in the amount of anomerisation (*Figure 6.10*) compared to that observed in the reactions of *ribo*-**53** (44% *vs* 8%, after 7 days at 40°C, 1M P_i, pH7), and show potential for significant anomerisation even without the presence of additional P_i (45% *vs* 7%, *arabino*-**53** and *ribo*-**53** respectively, after 7 days at 40°C, H_2O, pH7). This again confirms that there is a

significant thermodynamic preference for the *ribo*-**53** isomer over the *arabino*-**53** isomer.

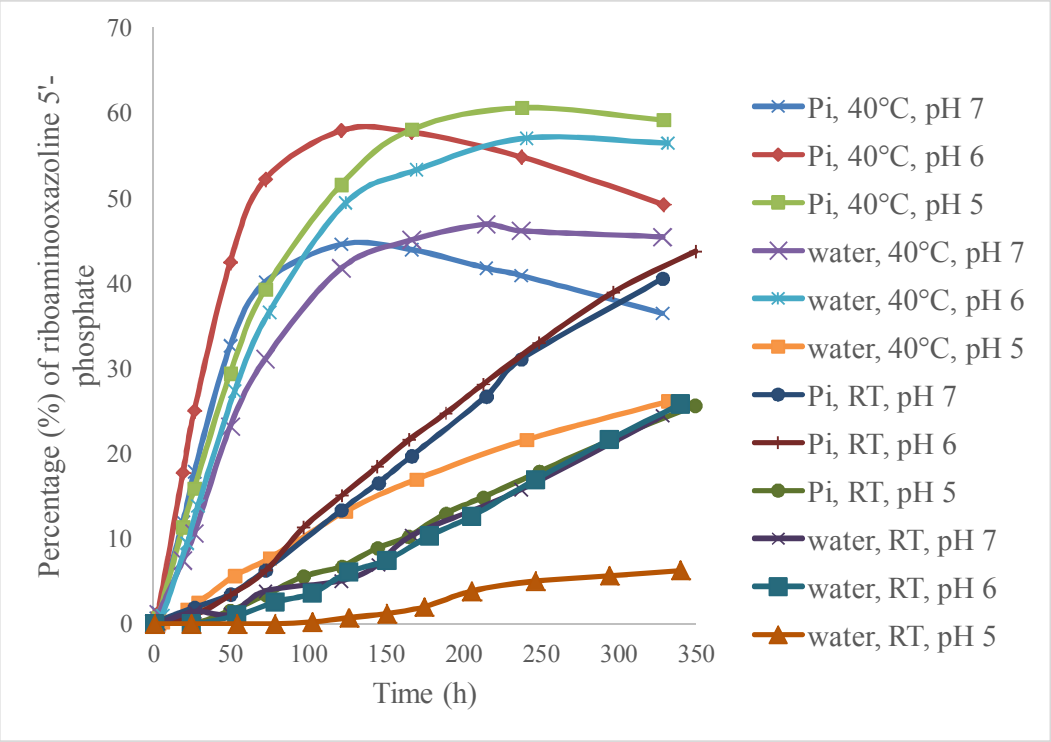


*Figure 6.10: Incubation of arabinoaminooxazoline-5'-phosphate (arabino-**53**; 0.1M) at various pH and temperatures, in H_2O/D_2O (1:1) or 1M P_i buffer (P_i) over time. Analysed by ^1H NMR spectroscopy. Percentages are reported as relative to identifiable species in solution.*

When comparing these results to the incubation of non-phosphorylated aminooxazolines *ribo*-**49** and *arabino*-**49**, we can see a clear increase in the chemistry observed in the **53** incubations, even without the presence of P_i buffer (*Table 6.4*, entries 1-2 and 5-6). Indeed in the incubation of *arabino*-**49** in the absence of P_i, there is no observed conversion to *ribo*-**49** after 7 days at 40°C (pH 7; *Table 6.4*, entry 5). Whereas in the same conditions in the incubation of 5'-phosphate *arabino*-**53**, there is 45% anomerisation to *ribo*-**53** (*Table 6.4*, entry 6). This indicates that the presence of a phosphate moeity in the aminooxazoline **53** acts as general acid/base catalyst and therefore, P_i is no longer a requirement.

Entry n°	Species	Conc. of P_i (M)	^{1}H NMR integration relative to other species integrated (%)					
			ribo-**49**/**53**	*arabino*-**49**/**53**	*ribo*-**106**/**107**	*arabino*-**106**/**107**	**98**b/ **100**b	**99**b/ **101**b
1	*ribo*-**49**	-	89.7	0	7.4	0	0.5	2.4
2	*ribo*-**53**	-	40	6.6	7.3	1.5	39.5	5.1
3	*ribo*-**49***[x]	1	54.4	6.7	7.9	0.2	17.3	2.9
4	*ribo*-**53**	1	38.1	4.6	20.7	3.2	30	3.4
5	*arabino*-**49**	-	73.9	0	10.6	0	1.1	14.5
6	*arabino*-**53**	-	10.8	45.1	5	3.4	32	3.7
7	*arabino*-**49***	1	9.6	27.4	17.8	19.1	10.2	15.7
8	*arabino*-**53**	1	6.9	44	9.1	11.3	25.6	3.3

*Table 6.4: Composition of species following the incubation of riboaminooxazoline (ribo-**49**), arabinoaminooxazoline (arabino-**49**), riboaminooxazoline 5′-phosphate (ribo-**53**) and arabinoaminooxazoline 5′-phosphate (arabino-**53**), at pH 7, after 7 days, at 40°C, with and without the presence of phosphate buffer.*As reported by Powner et al.; [x]5 days.*

The synthesis of *lyxo*-**53** from D-lyxose-5-phosphate (*lyxo*-**52**) and cyanamide (**4**) was attempted in similar conditions to the other **53**, however *lyxo*-**53** could not be isolated. We hypothesised that *lyxo*-**53** is particularly difficult to form selectively due to the steric clashes caused by the constituants all being on the same face (including the bulky and electron-dense phosphate group; *Section 4*, *Figure 4.11*). Indeed, it has been reported that non-phosphorylated *lyxo*-**49** is the only pentose aminooxazoline (**49**) to have a pronounced preference for the pyranosyl form (5:1 pyranosyl/furanosyl)[272], which is linked to the afforemention steric clashes. This pyranosyl form is inaccesible in the 5′-phosphate series of aminooxaxolines (**53**), due to the masked 5′-hydroxyl, thereby blocking the formation of the favoured isomer and any possible sugar ring closure. A study of the formation of *lyxo*-**53** from the free sugar *lyxo*-**52** with **4** (2 eq) at high concentration (1M, pH 7, 60°C, 3 h) showed the formation of an aromatic species (6.65 ppm, s, 1H; *Figure 6.11*) interpreted as **100**a, consistent with the hypothesis that sugar ring-closure is disfavoured in the conditions tested. Another study of a more dilute reaction

(22mM) showed the formation of *xylo*-**53** (alongside **100**a) after 3 h at 60°C (confirmed by spiking with an authentic sample, *Figure 6.12*).

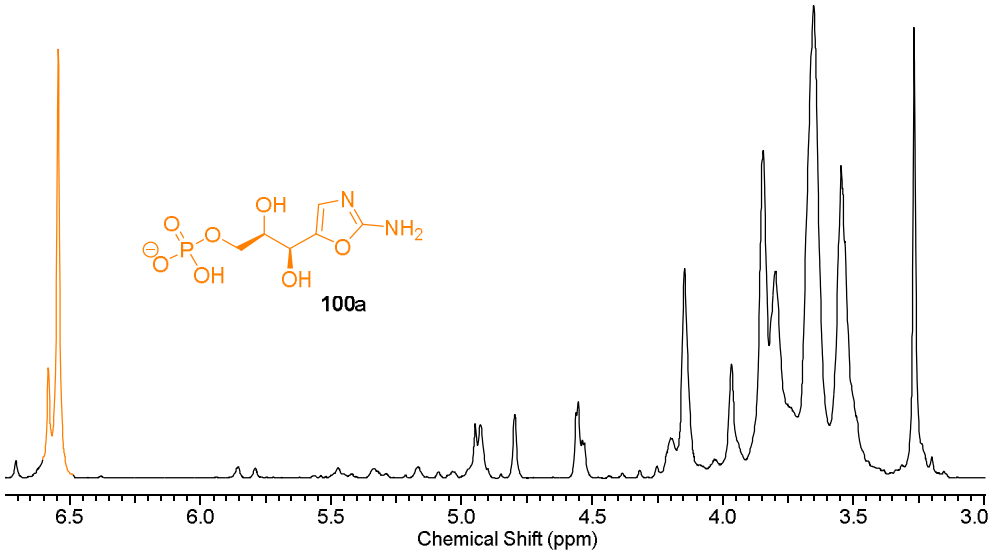


*Figure 6.11: ^{1}H NMR spectrum (600 MHz, H_2O/D_2O 9:1, 6.15-3.00 ppm) showing the formation of an aromatic species **100**a after 3 h of incubation of lyxose 5-phosphate (lyxo-**52**; 1M) with cyanamide (**4**; 2 eq), at pH 7 and 60°C. The characteristic aromatic peak for the **100**a has been highlighted.*

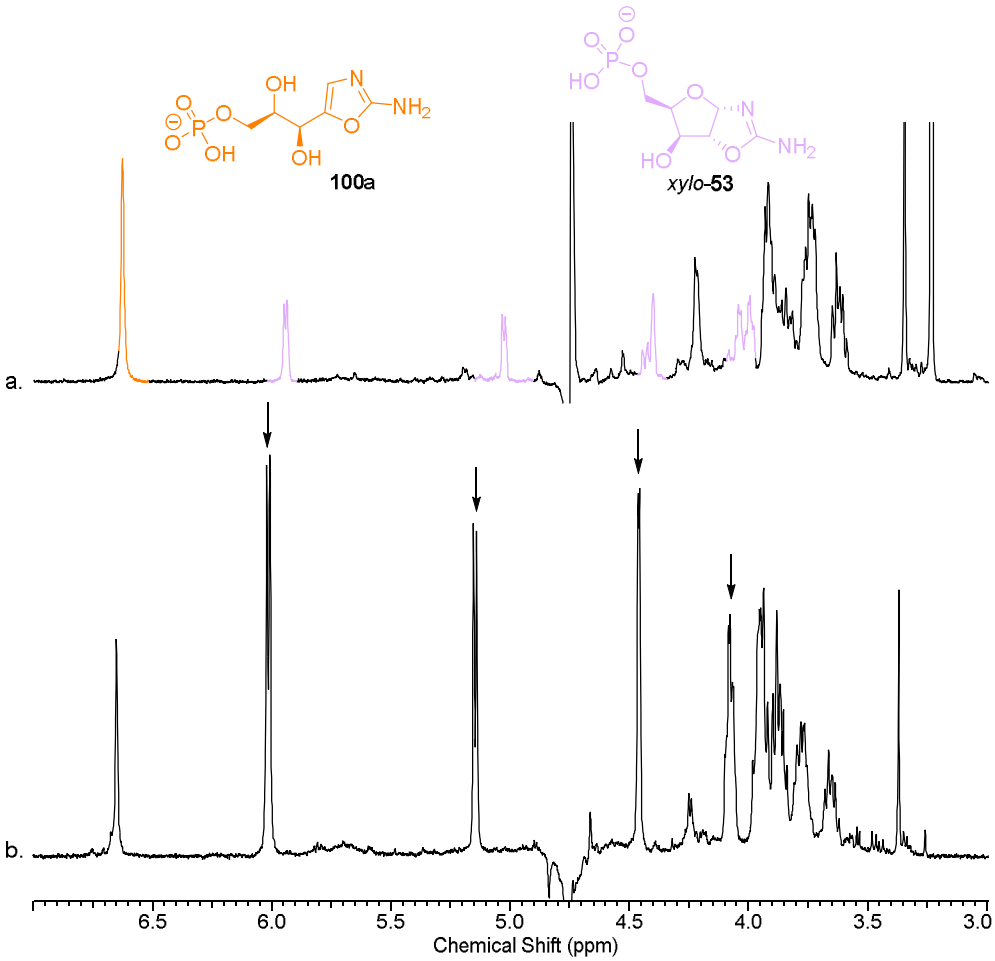
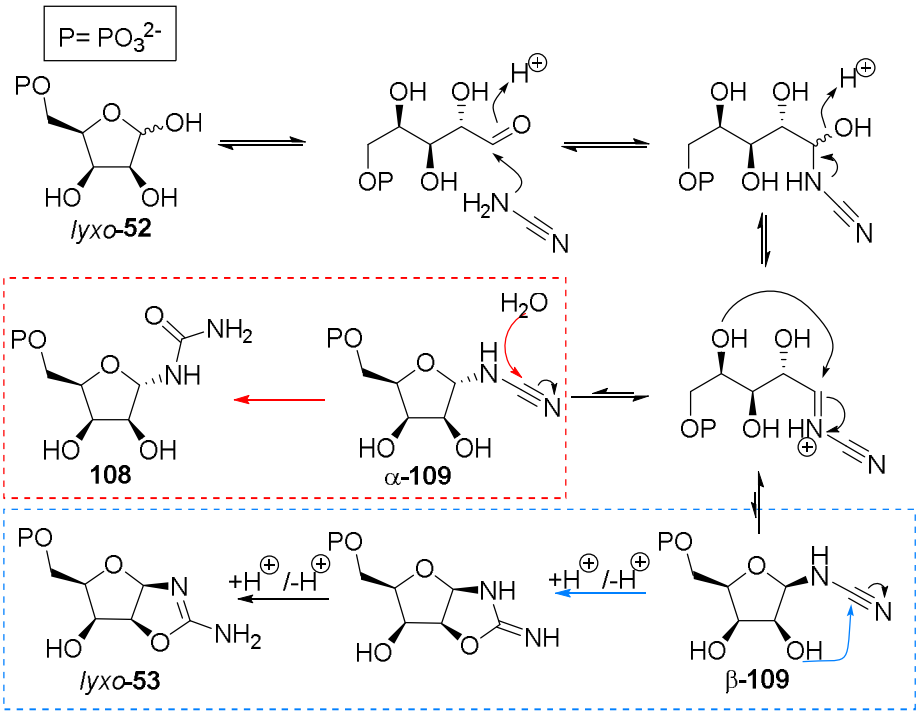


*Figure 6.12: ^{1}H NMR spectra (400 MHz, H_2O/D_2O 9:1, 7.00-3.00 ppm) showing the reaction of lyxose 5-phosphate (lyxo-**52**; 22mM) with cyanamide (**4**; 5 eq) in water at pH 7 after 3 h of heating at 60°C and subsequent signal amplification observed by spiking with a synthetically prepared authentic standard. a. Crude reaction ^{1}H NMR showing formation of aromatic species **100**a and xyloaminooazoline-5′-phosphate (xylo-**53**). b. Spiked with a synthetic sample of xylo-**53**. Characteristic peaks for the compounds have been highlighted.*

This confirmed that, as expected, *lyxo*-**53** is disfavoured at 60°C and the oxazoline-5′-phosphate species would preferentially sit as the aromatic form, **100**a, or cyclise to its more stable *xylo*- isomer, *xylo*-**53**. Analysis of the reaction of sugar phosphate *lyxo*-**52** (0.1M) with cyanamide (**4**; 2 eq, at pH 7) in the presence of ammonia (3% $NH_4OH_{(aq)}$) at RT revealed the formation of minor amounts of *lyxo*-**53** (15% of species in solution after 124 h, RT, based on ^{1}H NMR intergration), alongside the formation of hydrated open-chain *lyxo*-**101** and

a new compound **108** (60% of species in solution after 124 h; *Figure 6.13*, *Table 6.5*). We proposed that **108** could be the result of attack of **4** onto *lyxo*-**52** leading to adduct **109**, followed by hydrolysis of the cyanamide moiety, instead of cyclisation (*Scheme 6.4*, red pathway). This could be due to the **4** attack predominantly leading to an α-adduct **109**. Due to the obligate *cis*- relationship between C1′ and C2′ in aminooxazolines, a preference for the α-anomer in **109** would increase the time preceeding the cyclisation needed to form the aminooxazoline ring and thus allow the opportunity for hydrolysis. The H-C1′ (4.88 ppm, d, J = 3.9 Hz) signal for **108** coupling by HMBC to ^{13}C resonance signal at 163.8 ppm corroborates this interpretation.

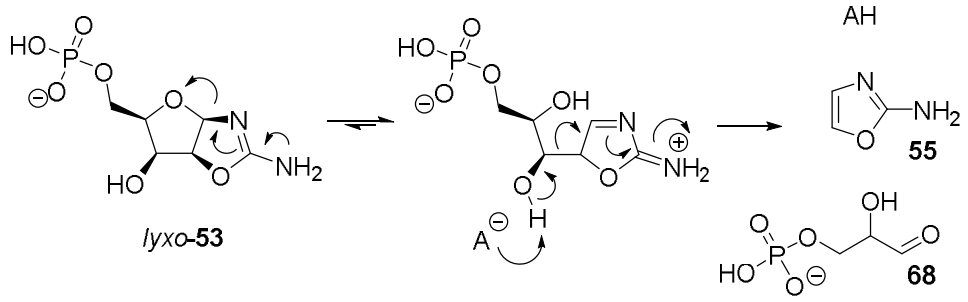


*Scheme 6.4: Proposed mechanism for the reaction of lyxose 5-phosphate (lyxo-**52**) with cyanamide (**4**) leading adduct **109**. Cyclisation of **109** to lyxoaminooxazoline 5′-phosphate (lyxo-**53**; blue pathway) and hydrolysis of the cyanamide moiety of **109** to **108** (red pathway).*

Time (h)	^{1}H NMR integration relative to other species integrated (%)					
	lyxo-**53**	*lyxo*-**52**	**108**	**100**b	**101**b	Other
0	6.4	8.9	19.1	0	65.6	0
2	6.6	8.6	24.3	0	60.5	0
8	7.4	8.1	29.6	0	54.1	0.7
13	8.3	7.5	32.3	0	50.6	1.4
25	10	8	40	0	40	2
43	12	8.4	46.8	0	30	2.8
100	14.7	8.8	58.8	1.5	13.2	2.9
124	15.4	9.2	60	1.5	10.8	3.1

*Table 6.5: Composition of species of the reaction of lyxose 5-phosphate (lyxo-**52**; 0.1M) with cyanamide (**4**), at pH 7, 3% $NH_4OH_{(aq)}$ at RT, over time.*

When the same reaction was performed at 60°C, we observed the additional formation of aromatic **100**a (35% after 15 h), as well as the appearance of 2-aminooxazole (**55**) and methylglyoxal (**96** ; 9% after 15 h). The latter is thought to be due to the retroaldol reaction of *lyxo*-**53** (*Scheme 6.5*) into oxazole **55** and glyceraldehyde-3-phosphate (**68**) followed by E1cB elimination to **96** (as discussed in *Section 3*; *Figure 6.14*, *Table 6.6*).



*Scheme 6.5: Proposed mechanism for the observed retroaldol reaction of 5'-phosphate lyxoaminooxazoline (lyxo-**53**) to 2-aminooxazole (**55**) and glyceraldehyde-3-phosphate (**68**).*

Time (h)	^{1}H NMR integration relative to other species integrated (%)						
	lyxo-**53**	*lyxo*-**52**	**108**	**55**	**100**b	**101**b	Other
2	16.1	8.1	50	0	3.2	22.6	0
4	18.5	9.3	50	1.9	7.4	13	0
6	15.8	9.5	45.8	3.2	14.2	11.1	0.5
7	13.6	9.5	40.7	5.4	19	10.9	0.9
11	12.2	11	32.9	7.3	26.8	8.5	1.2
13	9.1	11.9	31.1	8.2	31.1	7.3	1.2
15	8.5	11.9	27.2	9.3	34.8	6.8	1.4

*Table 6.6: Composition of species of the reaction of lyxose 5-phosphate (lyxo-**52**; 0.1M) with cyanamide (**4**), at pH 7, 3% $NH_4OH_{(aq)}$ at 60°C, over time.*

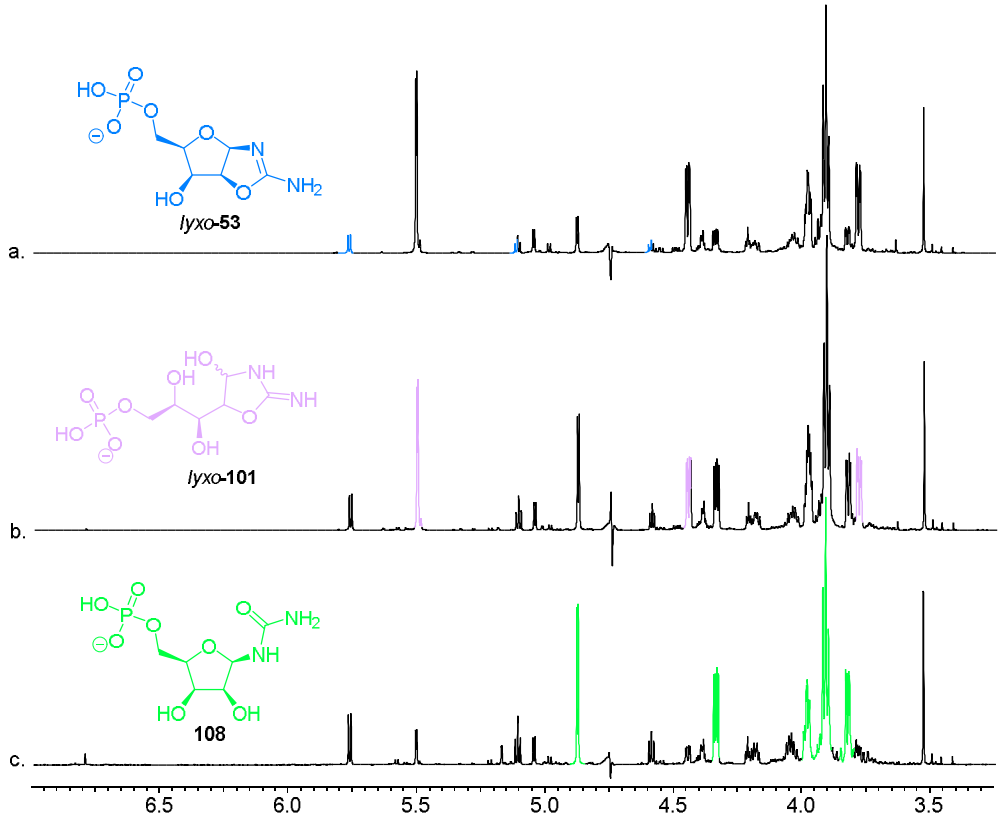


*Figure 6.13: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 7.00-3.25 ppm) showing the formation of a new species **108** (green), alongside lyxoaminooxazoline 5'-phosphate (lyxo-**53**; blue), and open-chain hydrate lyxo-**101** (pink), over time in the reaction of lyxose 5-phosphate (lyxo-**52**; 0.1M) with cyanamide (**4**; 2 eq), at pH 7, 3% $NH_4OH_{(aq)}$ at RT. Reaction after: a. 0.2 h; b. 25 h; c. 124 h. Characteristic peaks for each compound have been highlighted.*

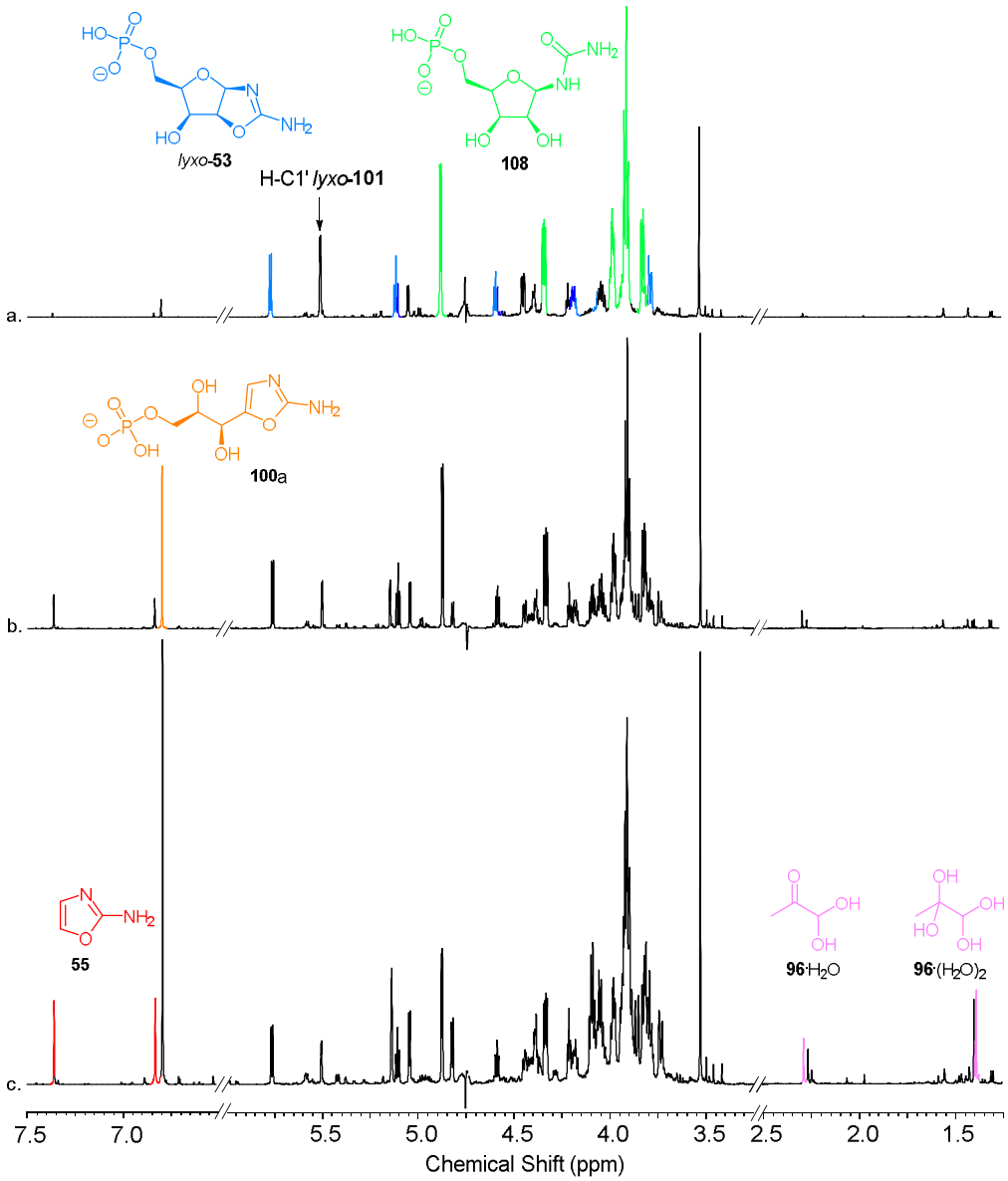


*Figure 6.14: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 7.50-1.25 ppm) showing evidence of retrosynthesis (formation of 2-aminooxazole (**55**) and methylglyoxal (**96**) over time) alongside lyxoaminooxazoline-5'-phosphate (lyxo-**53**) formation and aromatic species **100**a, in the reaction of lyxose 5-phosphate (lyxo-**52**; 0.1M) with cyanamide (**4**; 2 eq), at pH 7, 3% $NH_4OH_{(aq)}$ at 60°C. Reaction after: a. 2 h; b. 7.5 h; c. 15 h. Characteristic peaks for the compounds have been highlighted.*

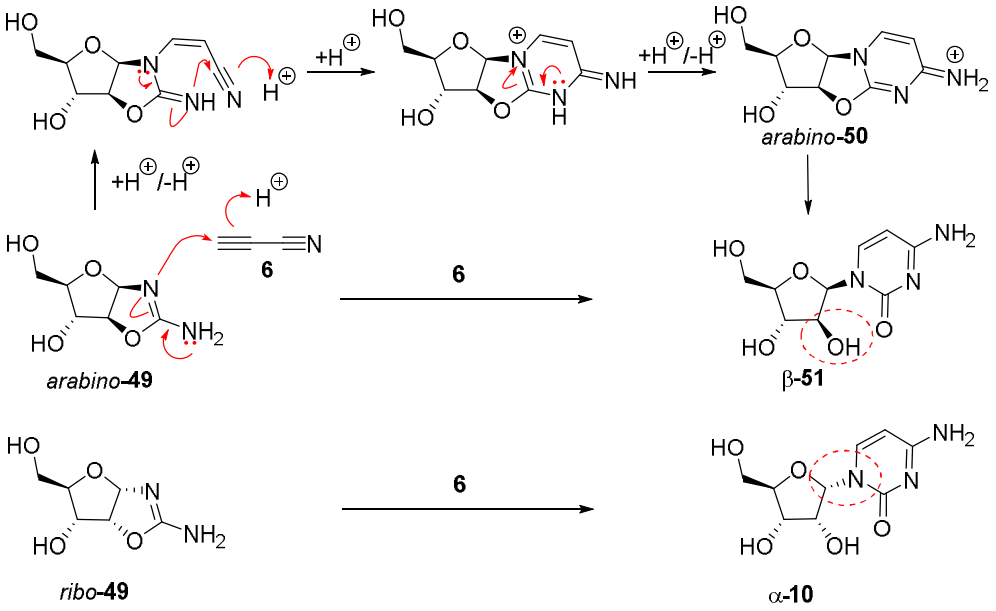
Even though we could not carry out anomerisation experiments with *lyxo*-**53**, it was clear from these results that the cyclised aminooxazoline *lyxo*-**53** was disfavoured at both RT and 60°C. This was consistent with the observed lack of

lyxo-**53** in our muticomponent reaction of glycidaldehyde (**69**), 2-aminooxazole (**55**) and P_i discussed in *Section 4*. We proposed that there is increased steric hindrance (compared to the other **53** isomers) due to the electron-dense phosphate group on the 5′ and all the other constitutents being on the same face, thereby preventing ring-closure (see *Figure 4.11*, p. 135). This in turn could explain the decreased reactivity observed in the anomerisation experiments of *xylo*-**53** compared to *ribo*-**53**, as the increased stability of *xylo*-**53** over *lyxo*-**53** would change the proportion of the mixture of compounds in favour of *xylo*-**53** (eg: 40% and 38% *ribo*-**53** after 7 days of *ribo*-**53** incubation in the absence and presence of P_i respectively *vs* 57% and 44% *xylo*-**53** after 7 days of *xylo*-**53** incubation in the absence and presence of P_i respectively; at 40°C, pH 7; *Tables 6.1/6.2*, entries 3-4).

In conclusion, we proposed that due to the presence of the 5′-phosphate, **53** can anomerise in the absence of excess P_i and in mild conditions (<40°C), leading to a simplification of a mixture of the four diastereomers of **53** in favour of *ribo*-**53** and *xylo*-**53**. In the context of the multicomponent reaction mentioned above (and discussed in *Section 4*), where we obtain a mixture of *ribo*-**53**, *arabino*-**53** and *xylo*-**53** (1:0.3:0.9; *lyxo*-**53** was not observed in its cyclic form and was shown to be sitting as the aromatic form **100**a), we can conclude that over time we would be left with only 2 diastereomers: *ribo*-**53** and *xylo*-**53**. This simplification brings us closer to a comprehensive route to the selective synthesis of canonical pyrimidine ribonucleotides: β-ribocytidine-5′-phosphate (β-**40**) and β-ribouridine-5′-phosphate (β-**110**).

6.2. Phosphoryl transfer

Sanchez and Orgel previously showed that reacting aminooxazolines *arabino*-**49** or *ribo*-**49** with cyanoacetylene (**6**)[xviii] yielded β-arabinocytidine (β-**51**; 10%) and α-ribocytidine (α-**10**; 10-20%) respectively[270] (both of which are one inversion away from the canonical β-ribocytidine (β-**10**), in the C2′ and C1′ position respectively, *Scheme 6.6*).

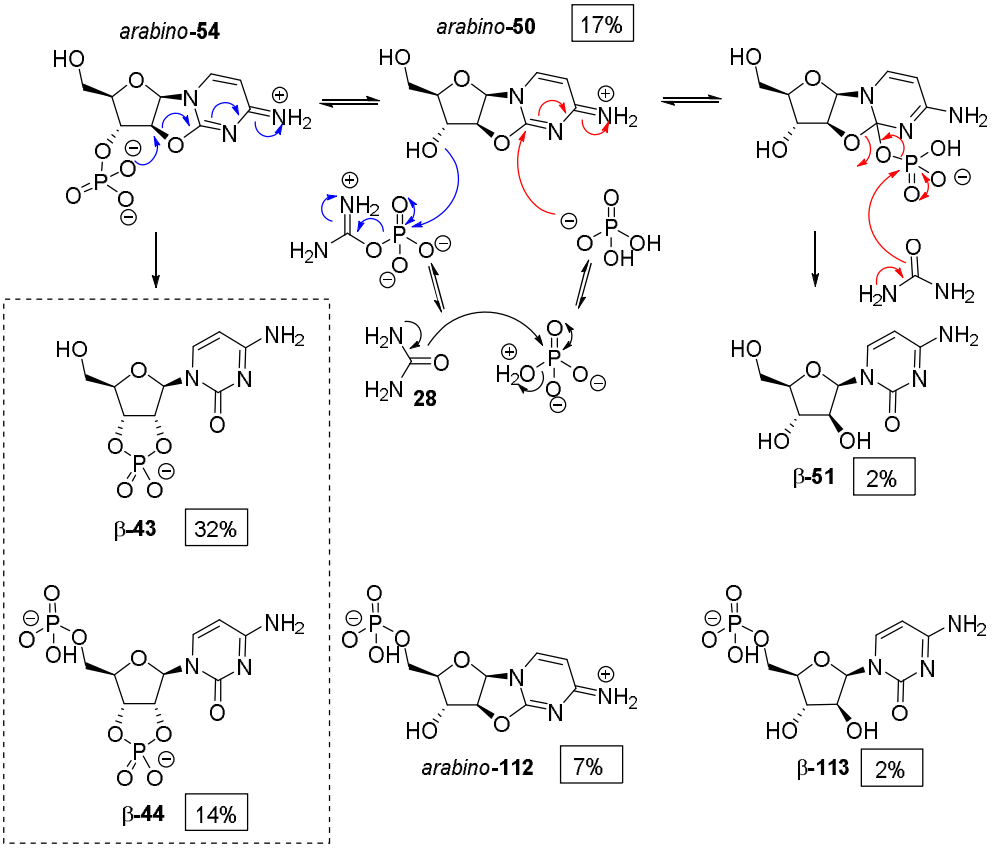


*Scheme 6.6: Mechanism of the reaction of aminooxazolines (**49**) with cyanoacetylene (**6**) (as demonstrated on arabino-**49**), leading to the formation of anhydrocytidine **50**. Subsequent hydrolysis was observed by Sanchez and Orgel[270] in their studies in water, without the use of P_i buffer. Stereochemical differences with the canonical β-ribocytidine (β-**10**) are denoted by a red circle.*

The Sutherland group latter demonstrated that this synthesis was achieved *via* the anhydronucleoside, 2,2′-anhydrocytidine (**50**), and the subsequent hydrolysis observed was due to a rise in the pH of the reaction[33]. They demonstrated that this

[xviii] Sanchez and Orgel's reactions were done as one-pot starting with 100mM ribose (**9**/)arabinose (**13**) along with 2 eq cyanamide (**4**) in 30% aqueous ammonia, heated at 100°C for 15 min, followed by addition of 3 eq of cyanoacetylene (**6**) and continued heating for 1 to 5 h.

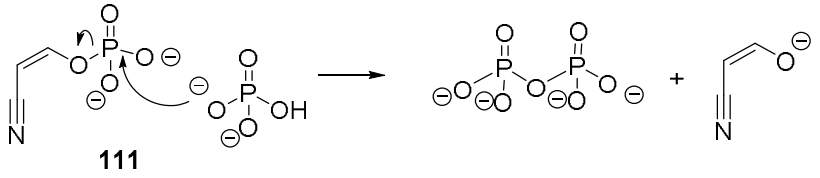
hydrolysis can be avoided by using P_i buffer to keep the pH around the optimum pH 6.5. Another advantage attributed to the P_i buffer in this reaction is its potential to also act as a chemical buffer as any excess of **6** can be quenched by the P_i (forming cyanovinyl phosphate **111**, *vide infra*) limiting the possibility of further unwanted reactions with the hydroxyl groups of the anhydronucleosides formed. Furthermore, the new-found action of the P_i buffer in this reaction enabled them to cement isolated anhydronucleosides as an essential compound *en route* to the canonical nucleotides, accessible *via* prebiotic reactions. Indeed, it was now possible to perform a phosphorylation-mediated rearrangement[271,312] *via* dry-state phosphorylation of *arabino*-**50** at the C3′ position, which is followed by cyclisation onto the C2′, leading to stereochemical inversion at the C2′ and finally yielding the canonical β-ribocytidine-2',3'-cyclic phosphate (β-**43**) in good yield (46% - including β-ribocytidine-2',3'-cyclic-5′-bisphosphate (β-**44**); the remainder of the compounds are notably: 17% un-reacted starting material - *arabino*-**50**; 7% 5′-phosphorylated *arabino*-**50** - *arabino*-**112**, and 4% minor hydrolysis products: β-**51** and its 5′-phosphate version – β-**113**; *Scheme 6.7*)[33]. This high yielding phosphorylation was achieved by drying a solution of anhydrocytidine **50** onto glass-fibre discs (40°C for 2 days), in the presence of ammonium chloride (2 eq), urea (**28**; 5 eq) and pyrophosphate (PP_i; 0.5 eq) and heating these disks at 100°C for 24 h.



Scheme 6.7: Results of the dry-state urea-mediated phosphorylation reported by Powner et al.[33] *on arabino-anhydrocytidine arabino-**50**. Mechanisms are shown for the activation of phosphate by urea (**28**), phosphate-mediated hydrolysis (red) and C3'-phosphorylation (blue). Yields of the products obtained are given (%) and canonical products are shown in the dashed box.*

Dry-state phosphorylations had already been demonstrated by Lohrmann and Orgel, who had achieved an un-selective but efficient phosphorylation (>90%) of nucleosides in the presence of urea (**28**) in a few hours[225]. However, they did not observe high yielding phosphorylation when the same experiment was perfomed without the presence of **28**. Powner *et al.*[33] reasoned that **28** helps the efficiency of these dry-state phosphorylations by first nucleophilically attacking the phosphate, thereby acting as a good leaving group when the phosphate is attacked by another hydroxyl group (such as C3′ OH of anhydrocytidine **50** in their case, *Scheme 6.7*, *vide supra*). Furthermore, Powner *et al.* pointed out that **28** can be formed as a by-product during an earlier stage of their **50** synthesis (in the reaction of glycolaldehyde (**3**) and cyanamide (**4**) to form 2-aminooxazole (**55**), urea (**28**) is

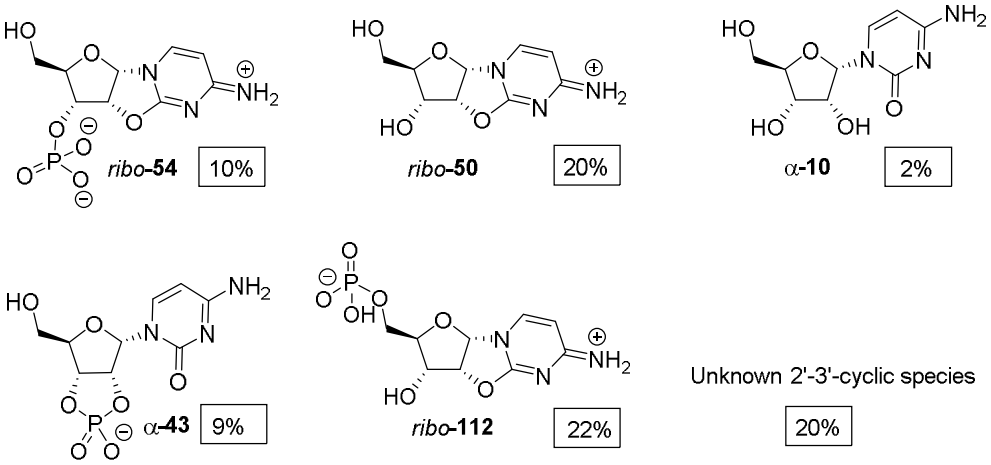
formed when there is an initial excess of **4**), thereby reinforcing its use at this latter stage of their synthesis. Ammonium chloride is used in these reactions in order to limit the amount of carbamylation that can occur alongside phosphorylation by inhibiting the decomposition of **28**[225]. Beck *et al.*[313] first made the case for the prebiotic use of ammonium phosphate ($NH_4H_2PO_4$) as a dry-state phosphorylating salt, as it has the advantage of being able to be formed in neutral or alkaline environment (through the condensation of ammonia and phosphate), while driving the pH of the environment down (where dry-state phosphorylation was observed to occur more readily) by virtue of the evaporation of ammonia in the event of desiccated heating. Powner *et al.* used both this inorganic phosphate $NH_4H_2PO_4$ and pyrophosphate (PP_i) in separate dry-state phosphorylations[33]. They argued for the prebiotic use of PP_i, which can be obtained by reaction of cyanovinyl phosphate (**111**) and P_i[314] (**111** is obtained as a by-product of the aforementioned synthesis of **50** from aminooxazolines (**49**) and cyanoacetylene (**6**), where excess **6** is quenched by the P_i buffer, *Scheme 6.8*), thereby extending the range of possible compounds used for phosphorylation.



*Scheme 6.8: Mechanism for the formation of PP_i from P_i and cyanovinyl phosphate **111**[314]. **111** can be obtained prebiotically as a by-product of the cyanovinylation of aminooxazolines **49/53** in P_i buffer*[33].

In the case of the reactions performed on anhydrocytidine *arabino*-**50** by the Sutherland group[33], they observed a highly selective conversion of *arabino*-**50** to β-ribocytidine-2',3'-cyclic phosphate (β-**43**) thought to be due to initial phosphorylation on the C3′-OH being preferred over the C5′-OH. This preference was attributed to the conformation of *arabino*-**50** (as determined by X-ray crystallography) where the C4′ *endo*-pucker coupled with the n→π* interaction of *O*-C5′ and C2 orbitals[315] (*Section 3*, *Figure 3.2*) causes steric crowding and a decreased nucleophilicity of *O*-C5′, leading to *O*-C3′ being comparatively more

available for phosphorylation, in spite of the steric hindrance intrinsic to secondary alcohols (compared to a primary alcohol such as *O*-C5′). Following from the selective C3′ phosphorylation, a cyclic phosphodiester β-**43** can be formed through intramolecular nucleophilic substitution[271] because of the *trans*-relationship between the hydroxyl groups at C3′ and C2′ in *arabino*-**50** (*Scheme 6.7*, *vide supra*). This substitution is not possible in the *ribo*- configuration *ribo*-**50** (due to the *cis*- relationship between C3′ and C2′), which is reflected in the poor results obtained by Powner *et al.*[33] (29% non-canonical 2′-3′ cyclic species - α-ribocytidine-2',3'-cyclic phosphate (α-**43**) and 2 unknown species, *Scheme 6.9*). The cyclisation observed in the *ribo*- series is most likely due to attack of the phosphate onto the 2′ or 3′ of the hydrolysis product cytidine α-**10** (or on the 3′ of the anhydronucleoside and subsequent hydrolysis), followed by a second urea-activation of the phosphate. At this point the second free *cis*-hydroxyl group (2′ or 3′) is free to nucleophilically attack, leading to α-**43**.



*Scheme 6.9: Results of the dry-state urea-mediated phosphorylation as reported by Powner et al.[33] on ribo-anhydrocytidine (ribo-**50**). Yields of the products obtained are given (%).*

The prebiotic synthesis of 5′-phosphate anhydronucleotides **112** was previously shown by the Sutherland group to be possible in the same conditions as anhydrocytidine **50**[33], starting from 5′-phosphate aminooxazolines (**53**). As such we synthesised, in the same prebiotic conditions as reported (49mM **53**, 5.5 eq cyanoacetylene (**6**), 1.1 eq P_i, pH 6.5, RT), anhydrocytidines 5′-phosphate

ribo-**112**, *arabino*-**112** and *xylo*-**112** (starting with the corresponding **53**). These syntheses afforded the desired product selectively and efficiently (94% *ribo*-**112**, 95% *arabino*-**112** and 95% *xylo*-**112**, respective yields after 16 h). Once these were purified (by ion-exchange chromatography), we sought to investigate the results of dry-state phosphorylations and dry-state phosphoryl transfer reactions on these species. We hypothesised that in the case of **112**, the 5′-phosphate moiety would replace the need for inorganic phosphate to be present by inter-molecular or intra-molecular phosphoryl transfer. Furthermore, the lack of P_i would limit unwanted base hydrolysis (as depicted in the red pathway, *Scheme 6.7*, *vide supra*) due to the distal position of the tethered phosphate compared to the C2 atom. These studies were performed using the protocol optimised by Sutherland[33], with and without the addition of P_i[xix], in the presence of urea (**28**; 5-20 eq) and ammonium chloride (2 eq)[xx].

When following the Sutherland procedure that yielded the highest conversion from *arabino*-**50** to cytidine β-**43** on our 5′-phosphorylated compound *arabino*-**112** (5 eq urea (**28**), 2 eq ammonium chloride and 0.5 eq PP_i), we observed a similar proportion of 2′,3′ cyclic product (48%; cf. 46% observed in the phosphorylation of *arabino*-**50**) by ^{1}H NMR analysis (*Table 6.7*, entries 1-2). Two sets of 2′,3′ cyclic peaks were observed as overlapping multiplets in the 5.2-4.9 ppm region of ^{1}H NMR correlating to significantly downfield shifted 23-21 ppm phosphorous signal by ^{1}H-^{31}P HMBC, characteristic of a 2′,3′ cyclic phosphate. Our product was mostly bisphosphorylated as the major set of peaks are consistent with literature data for β-ribocytidine-2′,3′-cyclic 5′-bisphosphate (β-**44**)[33]. A smaller signal for an ABX system (4.0-3.8 ppm) was attributed to C5′-OH compounds **43** (which could be obtained through phosphoryl-transfer, a point to be returned to later). A similar reaction with $NH_4H_2PO_4$ (1 eq; 10 eq **28**) showed an equally high rate of conversion to 2′,3′ cyclic products (50%; *Table 6.7*, entry 3), again mostly bisphosphorylated. The presence of minimal 2′,3′ cyclic product β-**43** was confirmed by spiking with an authentic standard of

[xix] Ammonium dihydrogen phosphate (1 eq) or pyrophosphate (0.5 eq).

[xx] The resulting solution was dried onto glass-fibre discs (40°C), and subsequently heated to 100°C for 24 h. Analysis was done according to the published procedure.

β-ribocytidine-2′,3′-cyclic phosphate (β-**43**; *Figure 6.15, c.*). These results suggest that a similar mechanism to the one for *arabino*-**50** was at play (*Scheme 6.7*, *vide supra*), resulting in incorporating an additional phosphate into our molecule at the C3′ position, with subsequent intra-molecular rearrangement to yield bis-phosphorylated cytidine β-**44**.

Entry n°	Species phosphorylated	Phosphate	Urea (eq)	^{1}H NMR integration relative to other species integrated (%)		
				Anhydro-	2′,3′ cyclic	Other[a]
1	*arabino*-**50***	PP_i	5	24	46	30
2	*arabino*-**112**	PP_i	5	4	48	48
3	*arabino*-**112**	$NH_4H_2PO_4$	10	0	50	50
4	*ribo*-**50***	PP_i	5	52	29	19
5	*ribo*-**112**	PP_i	5	34	23	43
6	*ribo*-**112**	$NH_4H_2PO_4$	10	31	30	39

*Table 6.7: Composition of species following the dry-state phosphorylation of anhydronucleotides in the presence of ammonium chloride (2 eq), urea (**28**; 5-10 eq) and PP_i (0.5 eq) or $NH_4H_2PO_4$ (1 eq). *as reported by Powner et al.*[316] *Anhydro- refers to species that retain the 2,2'-anhyonucleotide linkage, including all potentially phosphorylation states; 2',3' cyclic refers to the species where a 2',3' cyclic phosphate correlation is observed by ^{1}H-^{31}P-HMBC, including 5'-OH and 5'-phosphate; [a]Other refers to all other possible products including (non 2',3' cyclic phosphate) nucleosides/nucleotides, and nucleobase loss.*

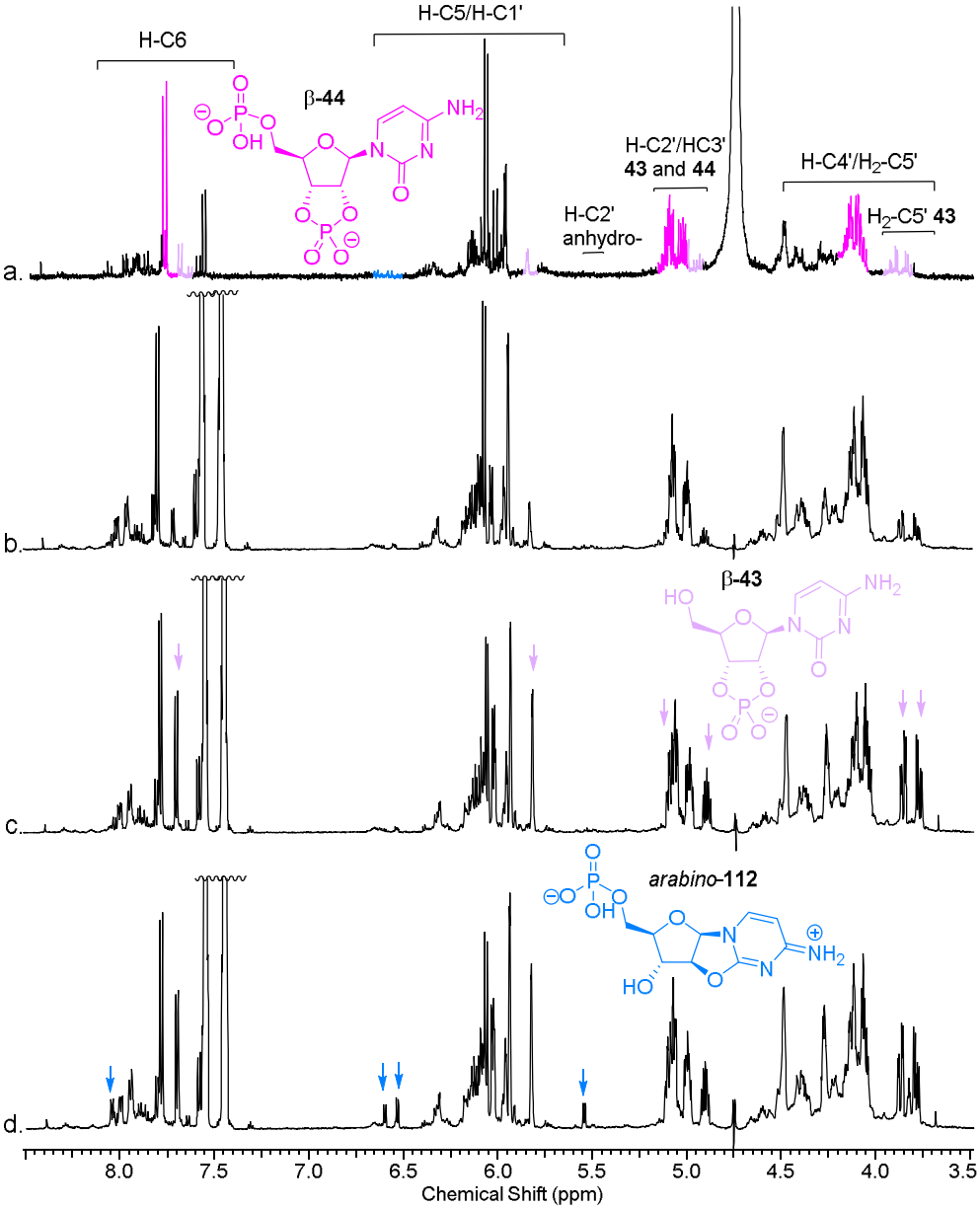
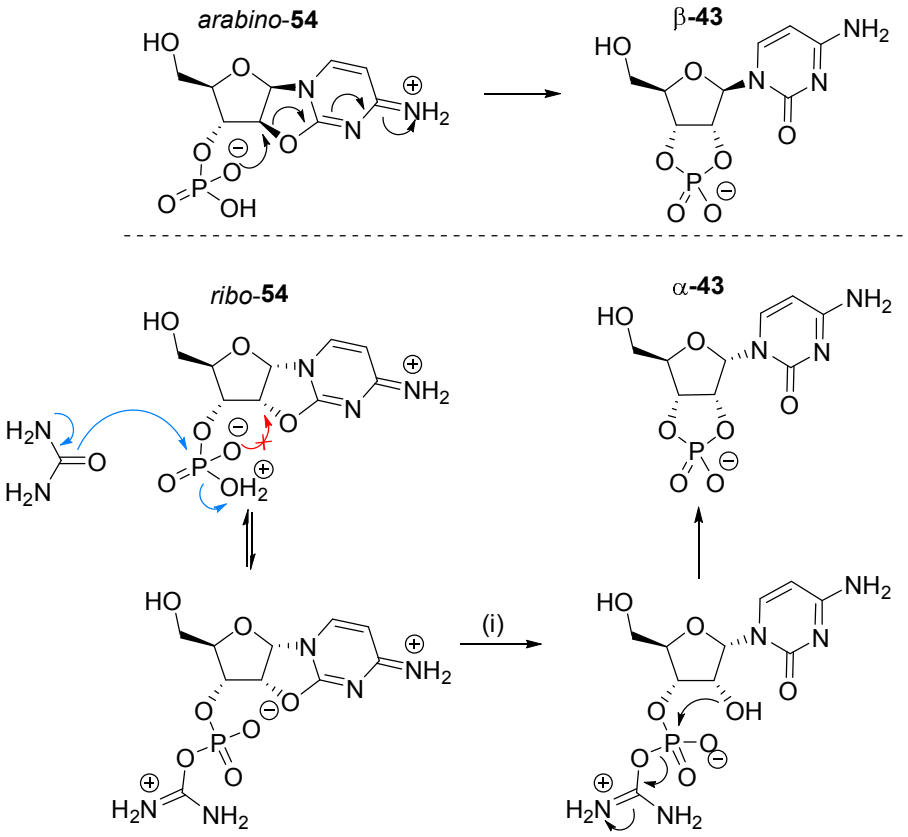


*Figure 6.15: ^{1}H NMR spectra (600 MHz, D_2O, 8.50-3.50 ppm) a. Arabinoanhydrocytidine 5'-phosphate (arabino-**112**), $NH_4H_2PO_4$ (1 eq) ammonium chloride (2 eq), urea (**28**;10 eq), in water, dried onto glass-fibre discs at 40°C (48 h), followed by heating at 100°C (24 h). NMR acquired after resuspension in D_2O, showing formation of β-ribocytidine-2',3'cyclic 5'-bisphosphate (β-**44**) and β-ribocytidine-2',3'cyclic phosphate (β-**43**). Characteristic peaks for the compounds have been highlighted and regions of interest are annotated. b. Addition of internal standard potassium hydrogen phthalate, followed by spiking with authentic standards; c. spiked with β-**43**; d. spiked with arabino-**112**.*

Next we investigated the *ribo*- variant riboanhydrocytidine 5′-phosphate (*ribo*-**112**; 5 eq urea (**28**), 2 eq ammonium chloride and 0.5 eq PP_i), in order to compare our results with those reported in the literature. We observed a low conversion to 2′,3′ cyclic product (23%; based on ^{1}H NMR integration of the H-C6′, *Table 6.7*, entry 5). The results were comparable to those obtained with anhydrocytidine *ribo*-**50** where 29% conversion to 2′,3′ cyclic products is reported (*Table 6.7*, entry 4). In the dry-state phosphorylation of *ribo*-**112**, a new ^{1}H-^{31}P HMBC correlation between a phosphorus signal (-0.9 ppm) and the H-C3′ region (4.90-4.70 ppm) was observed, along with a new triplet near the H-C2′ of *ribo*-**112**: 5.72 ppm, J = 5.1 Hz. These resonances are consistent with literature data for riboanhydrocytidine-3′-phosphate (*ribo*-**54**)[33] and therefore were assigned as 3′-phosphate *ribo*-**54**, or 3′,5′-bisphosphate *ribo*-**114**. Comparable results were obtained when performing the dry-state phosphorylation with $NH_4H_2PO_4$ (1 eq; 10 eq **28**; *Table 6.7*, entry 6) with some conversion to 2′,3′ cyclic product (30%) and *ribo*-**54**/**114**[xxi].

These results confirmed that a higher yield of 2′,3′ cyclic products is obtained in *arabino*- dry-state phosphorylations compared to *ribo*- phosphorylations whether 5′-phosphorylated or not, indicating that a similar mechanism occurs in 5′-anhydronucleotides (**112**; *Scheme 6.10*). The *cis*- conformation in *ribo*- compounds means that subsequently to initial phosphorylation on a free hydroxyl *via* activation of phosphate by urea (**28**) (the 3′-OH in the case of *ribo*-**112**; *Scheme 6.10*), an additional activation of the phosphate is needed for a cyclisation to occur, if and only if, the 2′-OH has been "released" from the anhydro-ring by means of nucleophilic phosphate attack (as depicted in *Scheme 6.7*, *vide supra*). Under the conditions used, with an excess of **28** over P_i, we postulated urea-activation of phosphate would be favoured over nucleophilic phosphate, the latter of which is necessary for phosphate-catalysed hydrolysis (as depicted in the red pathway in *Scheme 6.7*, *vide supra*). This would disfavour the anhydro-nucleobase hydrolysis (*Scheme 6.10*, *step i*), making the *ribo*- series unavailable for 2′,3′ cyclic formation.

[xxi] It was not possible to determine if this species was bisphosphorylated or not.



*Scheme 6.10: Proposed mechanism for the intra-molecular rearrangement from arabinose anhydrocytidine 3′-phosphate (arabino-**54**) to yield β-ribocytidine-2′,3′-cyclic phosphate (β-**43**), a mechanism not allowed in ribose anhydrocytidine 3′-phosphate (ribo-**54**; red), where the observed product α-ribocytidine-2′,3′-cyclic phosphate (α-**43**) is produced via activation of the 3′-phosphate by urea (**28**) (blue), hydrolysis of the anhydrocytidine and cyclisation. Phosphate activation and hydrolysis can occur in reverse order (not depicted). (i) Phosphate-mediated hydrolysis as depicted in Scheme 6.7, vide supra.*

It is of interest to note that 5′-phosphorylated anhydrocytidine (**112**) could undergo phosphoryl transfer of the 5′-phosphate. Dry-state experiments have been carried out by Orgel and Lohrmann[225] on nucleotide β-ribouridine 5′-phosphate (β-**110**; with 1 eq urea (**28**), 1eq NH_4Cl, heating at 100°C after drying) without additional P_i present. Their results showed appreciable levels of transphosphorylation, where dissociation of a 5′-phosphate from one molecule led to bisphosphorylation in another (forming ~10% β-ribouridine-2′(3′)-phosphate,

18% β-ribocytidine-2',3'-cyclic-5′-bisphosphate (β-**44**), ~25% β-ribouridine (β-**56**) and 28% β-ribocytidine-2',3'-cyclic (β-**43**) after 24 h at 100°C). Therefore, we proposed that a similar pathway could occur in our 5′-phosphorylated anhydronucleotides **112**. Dissociative activation of the 5′-phosphate of one **112** "donor" molecule by urea (**28**) would give rise to loss of phosphate in the "donor" leading to formation of unphosphorylated anhydrocytidine (**50**) alongside an activated phosphate (alternatively this de-phosphorylation step could occur on a 2',3'-cyclic-5′-bisphosphate **44** giving rise to 2',3'-cyclic monophosphate **43**). The activated phosphate can then go on to phosphorylate another molecule. If it phosphorylates another **112**, it would give rise, transiently, to a 5′,3′ bisphosphorylated compound anhydrocytidine-3′,5′-bisphosphate (**114**), which could then cyclise onto the 2′-position in a similar manner to that discussed previously (in the case of *arabino*-**112**; *Scheme 6.10*, *vide supra*). If the activated phosphate encounters a de-phosphorylated anhydronucleotide **50**, this would lead to preferential 5′-phosphorylation initially, due to less steric hindrance being present in a primary alcohol, leading back to **112**.

Investigative dry-state phosphorylation experiments were therefore carried out, without additional P_i present. Initial experiments on *arabino*-**112** were carried out with 5 eq of urea (**28**)[xxii] and showed a substantial amount of 2′,3′ cyclic phosphate species formed (37% of discernible species by ^{1}H NMR; *Table 6.8*, entry 1; *Figure 6.16, a.*). From the NMR analysis (^{1}H-^{31}P HMBC and ^{1}H NMR), we could distinguish two sets of 2′,3′ cyclic phosphate species attributed to β-ribocytidine-2′,3′cyclic-5′-bisphosphate (β-**44**) and to β-ribocytidine-2′,3′-cyclic phosphate (β-**43**; the latter was confirmed by spiking with an authentic standard; *Figure 6.16, c.*). β-**44** and β-**43** were found to be in comparable amounts (1:1.05 β-**44**/β-**43** by ^{1}H NMR). A similar experiment with 10 eq of **28** showed an increase in the proportion of 2′,3′ cyclic phosphate species (60%; *Table 6.8*, entry 2), again with both β-**44** and β-**43** present, but with a higher ratio of mono-phosphate β-**43** (1:2). This suggested that more **28** increased the propensity

[xxii] The same conditions were used as for the reactions described above: **112** was dried onto glass-fibre discs for 48 h at 40°C, in the presence of ammonium chloride (2 eq), urea (**28**; amount stated), in the absence of phosphate; followed by heating at 100°C for 24 h.

for dissociative phosphoryl-transfer[xxiii], due to increased activation of the phosphate.

Entry n°	Species phosphorylated	Urea (eq)	^{1}H NMR integration relative to other species integrated (%)		
			Anhydro-	2′,3′ cyclic	Other[a]
1	*arabino*-**112**	5	36	37	27
2	*arabino*-**112**	10	9	60	32
3	*ribo*-**112**	5	55	18	27
4	*ribo*-**112**	10	68	14	18
5	*arabino*-/*ribo*-*	10	28	30	42

*Table 6.8: Composition of species following the dry-state phosphorylation of anhydronucleotides 5'-phosphate **112** in the presence of ammonium chloride (2 eq) and urea (**28**; 5-10 eq). *Mixed experiment with arabino-**112** and ribo-**112** (0.5 eq each). Anhydro- refers to species that retain the 2,2'-anhyonucleotide linkage, including all potentially phosphorylation states; 2',3' cyclic refers to the species where a 2',3' cyclic phosphate correlation is observed by ^{1}H-^{31}P-HMBC, including 5'-OH and 5'-phosphate; [a]Other refers to all other possible products including (non 2',3' cyclic phosphate) nucleosides/nucleotides and nucleobase loss.*

[xxiii] Experiments with 20 eq of urea (**28**) did not show any significant difference and as such we will concentrate on the 10 eq reactions to better compare with previous experiments.

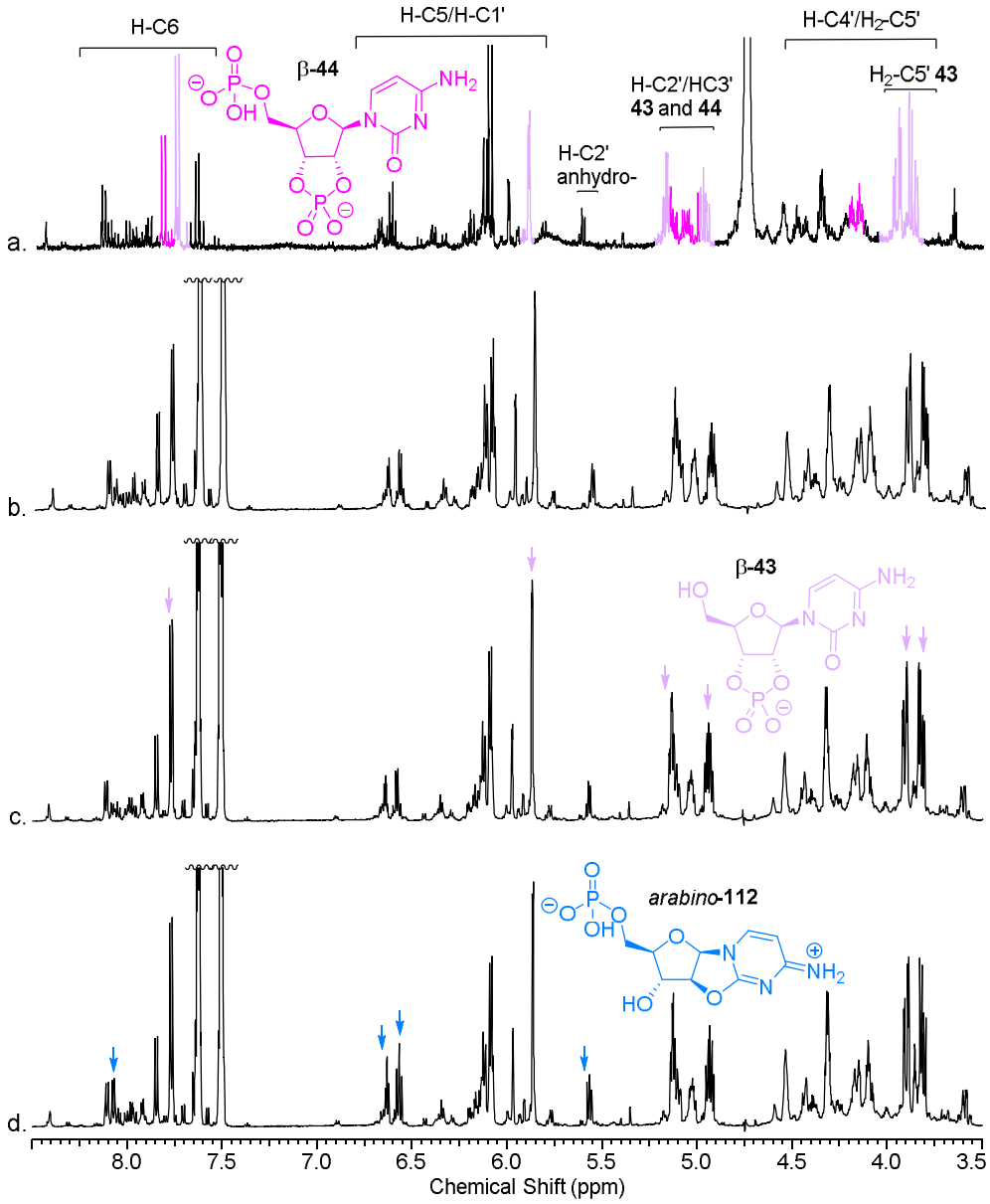


*Figure 6.16: ^{1}H NMR spectra (600 MHz, D_2O, 8.50-3.50 ppm) a. Arabinoanhydrocytidine 5'-phosphate (arabino-**112**; 240mM), ammonium chloride (2 eq), urea (**28**; 10 eq) in water, dried onto glass-fibre discs for 48 h at 40°C, followed by heating at 100°C for 24 h. NMR acquired after resuspension in D_2O, showing formation of β-ribocytidine-2',3'cyclic-5'-bisphosphate (β-**44**) and β-ribocytidine-2',3'cyclic phosphate (β-**43**). Characteristic peaks for the compounds have been highlighted and regions of interest are annotated. b. Addition of internal standard potassium hydrogen phthalate, followed by spiking with authentic standards; c. spiked with β-**43**; d. spiked with arabino-**112**.*

The same experiments were repeated with *ribo*-**112** and showed a marked decrease in 2′,3′ cyclic phosphate species formed compared to the *arabino*- series (18% and 14% with 5 eq and 10 eq of urea (**28**) respectively, based on ^{1}H NMR integration; *Table 6.8*, entries 3-4). The main compounds detected in these experiments were anhydronucleotides (55% and 68% with 5 eq and 10 eq of **28** respectively), again with the presence of a few peaks in the H-C2′ region of anhydronucleotides (5.90-5.60 ppm), suggesting various phosphorylation states: un-phosphorylated riboanhydrocytidine (*ribo*-**50**), mono-phosphorylated *ribo*-anhydrocytidine 5′-phosphate (*ribo*-**112**) or 3′-phosphate *ribo*-**54**, and bis-phosphorylated riboanhydrocytidine 3′,5′-bisphosphate (*ribo*-**114**), in accord with a dissociative transfer of phosphate.

It is of interest to note that increasing the proportion of urea (**28**) in this case led to a decrease in phosphate reactivity. This is consistent with the hypothesis that nucleophilic phosphate is needed for anhydro- hydrolysis to the nucleotide/nucleoside (*Scheme 6.7*, *vide supra*), an essential step *en route* to 2′,3′ cyclic phosphate formation in the *ribo*- series (*Scheme 6.10*, *vide supra*). Increasing the proportion of **28** two-fold therefore favoured electrophilic phosphate formation (phosphate-urea adduct), and further highlighted the mechanistic differences in the *arabino*- cyclisation and *ribo*- cyclisation (*Scheme 6.10*, *vide supra*). A mixed experiment (1:1 *ribo*-**112**/ *arabino*-**112**) showed that this divergence in reactivity between *ribo*-**112** and *arabino*-**112** can be observed *in situ* (*Table 6.8*, entry 5, *vide supra*; *Figure 6.17*). Indeed, most of the anhydrocytidine compounds were *ribo*- in configuration (determined by spiking with *ribo*-**112**; *Figure 6.17*, c.), while the majority of 2′,3′ cyclic species were β-**43**, suggesting they were formed from *arabino*-**112**.

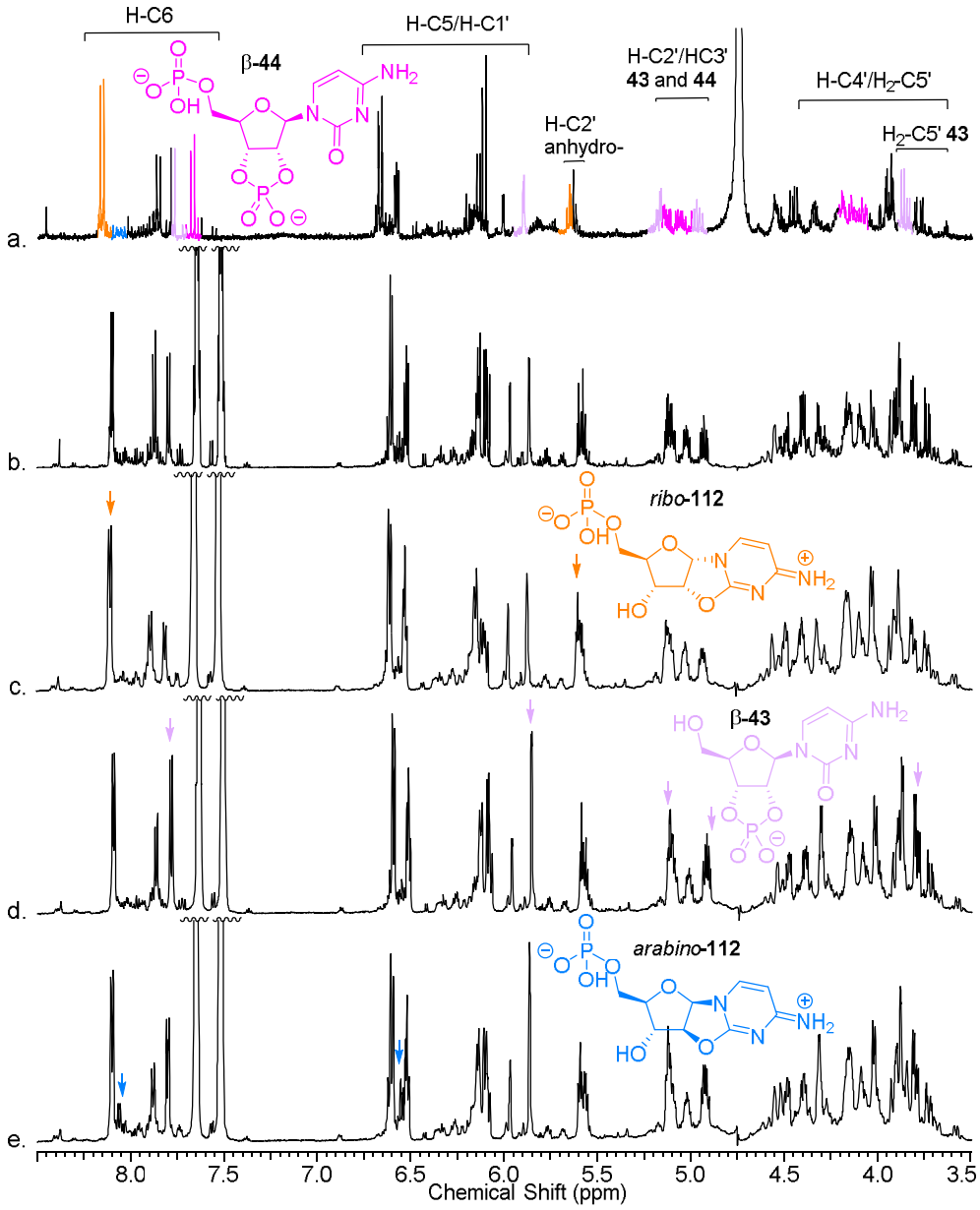


*Figure 6.17: ^{1}H NMR spectra (600 MHz, D_2O, 8.50-3.50 ppm) a. Arabinoanhydrocytidine 5'-phosphate (arabino-**112**) and ribo-**112** (0.5 eq each), ammonium chloride (2 eq), urea (**28**; 10 eq), in water, dried onto glass-fibre discs for 48 h at 40°C, followed by heating at 100°C for 24 h. NMR acquired after resuspension in D_2O, showing formation of β-ribocytidine-2',3'-cyclic 5'-bisphosphate (β-**44**) and β-ribocytidine-2',3'-cyclic phosphate (β-**43**), and retention of ribo-**112**. Characteristic peaks for the compounds have been highlighted and regions of interest are annotated. b. Addition of internal standard potassium hydrogen phthalate, followed by spiking with authentic standards; c. spiked with ribo-**112**; d. spiked with β-**43**; e. spiked with arabino-**112**.*

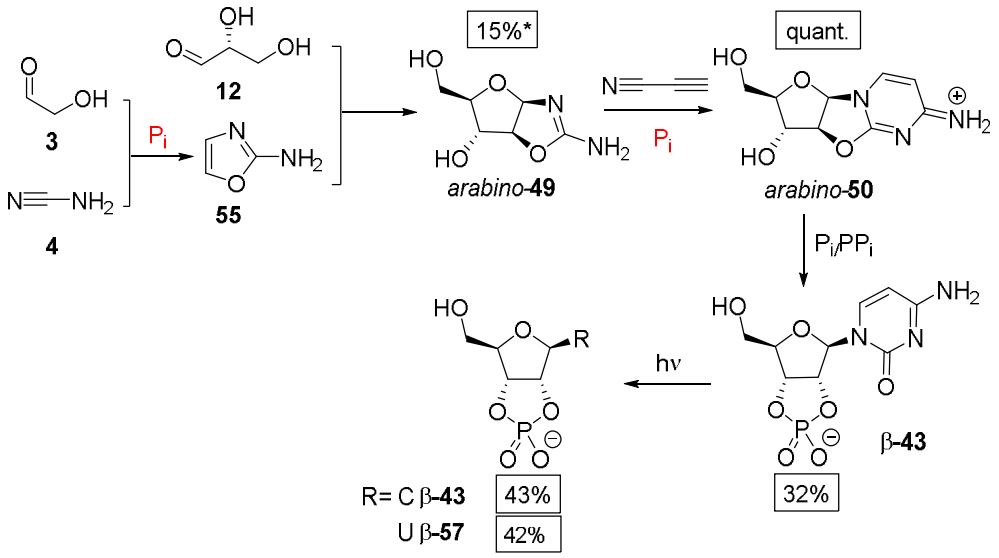
It was clear from these experiments that we were observing similar results to those obtained by Powner *et al.*[33], where arabinose anhydrocytidine (*arabino*-**50**), once phosphorylated on the 3′-position, can perform an intra-molecular rearrangement to yield the canonical cyclic phosphate β-**43** (or its bisphosphate β-**44**). Interestingly, this can be achieved without an excess of phosphate when starting with *arabino*-**112**, due to inter-molecular transfer replacing the need for an inorganic source of phosphate. We could imagine that in the event of a mixed synthesis of *ribo*-**112** and *arabino*-**112**, this phosphorylation could yield a way of both epimerising the *arabino*-**112** to the canonical β-**43** (linking into the synthesis established by the Sutherland group[316]) and retaining riboanhydrocytidine-5′-phosphate (*ribo*-**112**), that can be used for a parallel route that takes advantage of the 5′-phosphate present (*vide infra*).

These results would be most useful in the context of a selective prebiotic synthesis of *arabino*-**112** (which somehow avoided the efficient anomerisation from its precursor aminooxazolines 5'-phosphates *arabino*-**53** to *ribo*-**53**, *vide supra*). However, we concluded in the previous section that the multicomponent synthesis developed and discussed so far (of glycidaldehyde (**69**), 2-aminooxazole (**55**) and P_i, forming **53**; *Section 4*) led to *ribo*-**53** as the main diastereomer, which was later amplified by the facile anomerisation of *arabino*-**53** to *ribo*-**53** (*vide supra*), culminating in a *ribo*-**53** rich synthesis (with some *xylo*-**53**, a point to be returned to later).

In addition to the fact that dry-state phosphorylations would not be useful in our synthesis *via ribo*-**112**, one has to point out the advantages present from avoiding dry-state environments in prebiotic synthesis and from retaining the 5′-phosphate (as opposed to obtaining 2′,3′-cyclic phosphates). Firstly, while dry-state environments can be imagined in a prebiotic context (e.g. desert areas[225]), it has to be pointed out that all reactions leading up to the synthesis of anhydronucleotides are performed in aqueous solutions. The discrepancy between these two environments can be solved by referencing wet-dry cycles, or shore-line environments, where the amount of solvent would change over time. However,

this greatly limits the environments that can be used for prebiotic nucleotide synthesis, as well as the time-line of events: heating in the dry-state at the "wrong" point in the synthesis could destroy the precursors needed for anhydronucleotide/nucleotide synthesis. Furthermore, dry-state phosphorylations as performed by Powner *et al.*[316], are carried out on anhydronucleotide *arabino*-**50** in order to obtain the desired product β-ribocytidine-2′,3′-cyclic phosphate (β-**43**), ignoring the likely anomerisation of *arabino*-**50** to *ribo*-**50** (only 27% left after 7 days in 1M P_i buffer at 40°C; *Table 6.4*, entry 7, *vide supra*). Additionally, dry-state phosphorylations on *arabino*-**50** are unselective for the desired product β-ribocytidine-2′,3′-cyclic phosphate (β-**43**; 32%) with other chemical pathways competing with the desired one (other phosphorylations, 5′-phosphorylation of anhydrocytidine and of 2′,3′-cyclic species, and hydrolysis to β-arabinocytidine-(5′-phosphate)). This lack of selectivity is partially solved in the subsequent irradiation step where β-**43** is found to be stable to the irradiation conditions while other products resulting from the phosphorylations step are partially destroyed over the same time-frame. However, this synthesis leads to a significant loss of material over the four steps and thereby diminishes the effective concentration of β-**43** (only ~5% β-**43** and its uridine variant β-**57** combined[xxiv] over the 4-step synthesis from glyceraldehyde (**12**)/2-aminooxazole (**55**) to β-**43**/ β-**57**; *Scheme 6.11*).

[xxiv] Based on calculations carried out on reported yields for synthesis of aminooxazolines (**49**), the dry-state phosphorylations of *arabino*-**50** and *ribo*-**50**, and the irradiations described by Powner *et al*[316]. These calculations assumed a quantitative conversion of **49** to **50** and ignored the potentially anomerisation of *arabino*-**49** and *ribo*-**49**.



*Scheme 6.11: Four step prebiotic synthesis of activated pyrimidine ribonucleotides as described by Powner et al[33], showing the yields reported for each step leading to the natural stereochemistry (as depicted). Steps catalysed by P_i are shown in red. *Yield over two steps, based on glycolaldehyde (**3**).*

Secondly, the problem of oligomerisation should be raised. There is a need for regioselective control of oligomerisation of monomers in order to obtain the 3′-5′ phosphodiester linkage observed in naturally occurring RNA. The Szostak group proposed that some heterogeneity in the linkage could be beneficial to early RNA biopolymers by lowering the melting point of RNA duplex by having a mixture of 3′-5′ with 2′-5′, thereby facilitating the strand displacement needed for replication[317]. However, 2′-5′ phosphodiester linkages hydrolyse significantly faster than the 3′,5′ linkages (between two and three orders of magnitude faster)[233], suggesting the natural 3′-5′ linkage is necessary in the building of stable RNA oligomers.

Additionally, the 2′,3′-cyclic phosphate ring offers little activation towards a nucleophilic attack by a hydroxyl group[232] and any reactions done in water would also compete with the hydrolysis of the 2′,3′-cyclic phosphate to nucleotides 2′ and 3′-monophosphates[318]. The phosphate activation needed for ligation would then preferentially lead to re-cyclisation to form 2′,3′-cyclic phosphate, as the local concentration of the neighbouring hydroxyl group would be much higher

than any other nucleotide 5′-OH[223,236,237,319]. Some oligomerisations using 2′,3′-cyclic phosphates have been proposed such as dry-state oligomerisation (in the presence of catalysts, without a template) leading to short oligomers (<6 units)[320] with a mixture of 2′-5′ and 3′-5′ linkages; or the use of templates to aid ligation, where oligomerisations of nucleotides 2′,3′-cyclic phosphate leading preferentially to 2′-5′ phosphodiester linkages (95%) are formed at a rate comparable to the hydrolysis of the newly-formed bond[233,234]. The Sutherland group has proposed an alternative solution *via* a prebiotic protecting group strategy using acetylation. They found that acetylation of nucleotides 2′ and 3′-monophosphates leads to the preferential protection of 2′-OH of nucleotide 3′-monophosphates (in 52% yield within 24 h using thioacetate and cyanoacetylene (**6**), with minimal protection of the 3′-OH of nucleotide 2′-monophosphates)[319]. They found that this protecting group strategy increased the templated-ligation from 1% to 49%, while affording controlled selectivity to afford a 3′-5′ phosphodiester linkage. The acetyl group can then be removed prebiotically by a short incubation with ammonia[319]. However, the abundance of prebiotic activating agents leading to 2′ and 3′-monophosphates cyclisation[223,236,237,319], and the moderate yield of acetylation obtained in this synthesis, puts into question the prebiotic feasibility of this oligomerisation strategy.

In contrast, 5′-phosphate nucleotide oligomerisation is much less problematic. The use of variety of prebiotically plausible activation agents can lead to selective 5′-phosphate activation without the need for any other reactions to take place (such as acetylation) as no neighbouring hydroxyl group would lead to cyclisation[243,321]. Instead an activated 5′-phosphate can lead to effective ligation, with a preference for the canonical 3′-5′ phosphodiester linkage[244,245] of up to 80:1 (3′-5′/2′-5′ linkage respectively) when using 5′-triphosphates as the activated 5′-phosphate[243]. The non-enzymatic oligomerisation of 5′-triphosphates is a robust and efficient reaction[243,247] and is of particular interest to us considering the possibility of triphosphate incorporation into aminooxazolines using our epoxide/2-aminooxazole adduct **97** (a point to be returned to, *Section 7*).

In conclusion, the synthesis we have developed so far (of glycidaldehyde (**69**), 2-aminooxazole (**55**) and P_i, forming 5'-phosphate aminooxazolines (**53**); *Section 4*) allows for aqueous incorporation of phosphate selectively at the 5′-position, both allowing us to by-pass the need for problematic dry-state phosphorylations, and procuring us with the advantages inherent to 5′-phosphates. The markedly increased selectivity for the *ribo*- series obtained, in both the initial multicomponent reaction, as well as in its increased stabilisation (seen in the anomerisation studies, *vide supra*) pushed us to explore the intriguing idea of building nucleotides *via* 5′-phosphorylated *ribo*-compounds.

6.3. Anomerisation by irradiation

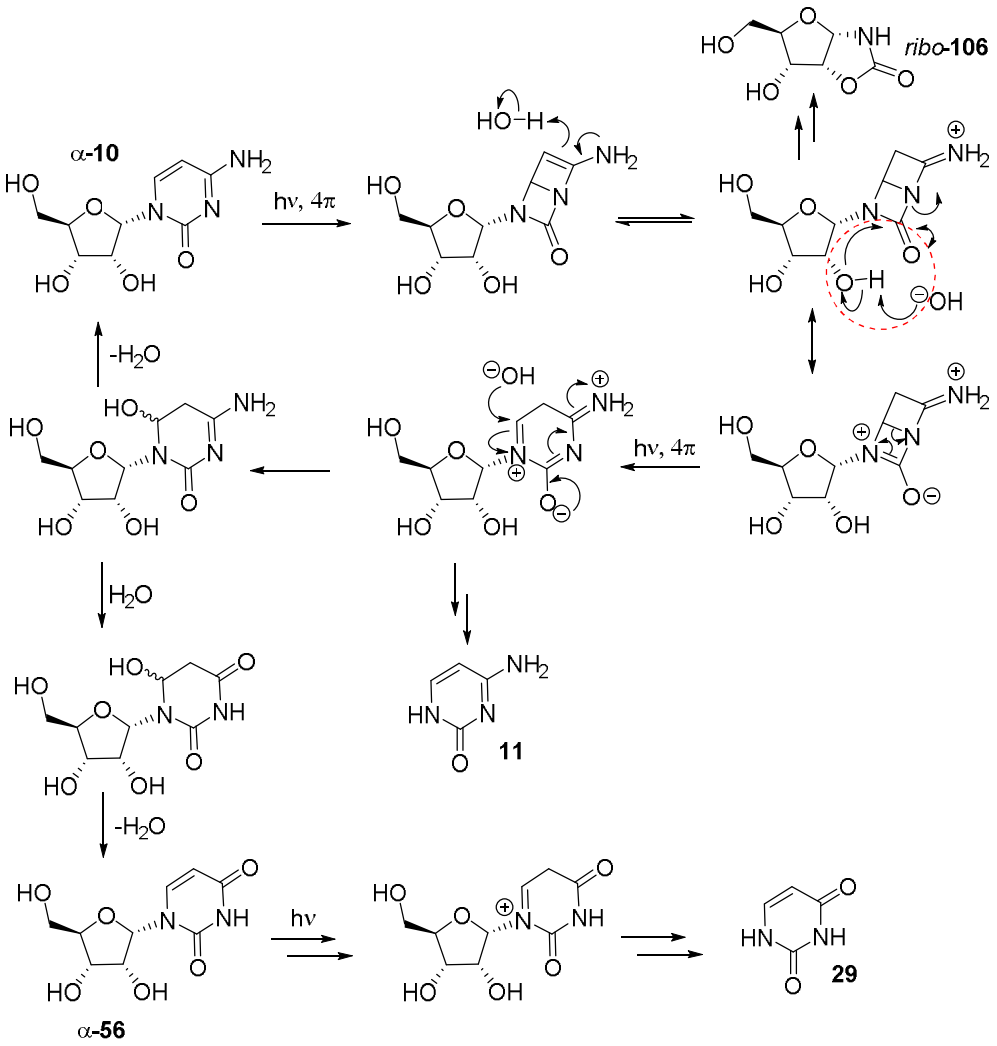
As discussed above, the use of anhydronucleotide compounds in our *ribo*-selective route was not fruitful, therefore leading us to explore other prebiotic avenues that could complete our synthesis to afford the canonical pyrimidine ribonucleotides 5′-phosphate: β-ribocytidine 5'-phosphate (β-**40**) and β-ribouridine 5'-phosphate (β-**110**).

Hydrolysis of anhydronucleotide compound *ribo*-**112** to the corresponding nucleotide α-ribocytidine 5′-phosphate (α-**40**) is an efficient reaction[322] but gives rise to unnatural stereochemistry in the C1′ position. Therefore, we considered precedents for prebiotically plausible C1′ anomerisation from α-*ribo* nucleotide α-**40** to the canonical β-**40**. Remarkably, Sanchez and Orgel first reported α to β anomerisation in *ribo*-cytidine (α-**10**) and *ribo*-cytidine 5′-phosphate (α-**40**)[270]. They reported low conversions of α-**10** to β-**10**, in the order or 4% after 6 h irradiation, and similarly low conversion yield of 6% α-**40** to β-**40** after 16 h irradiation. However, lack of details concerning by-products obtained from the mixture and substantial recovery of starting material: 53% and 32% respectively; along with the long work-up procedure in which compounds of interest could be lost (involving electrophoresis, chromatography, dephosphorylation and finally ion-exchange chromatography followed by analysis by UV spectra), led Powner *et al*.[323] to reassess these reactions.

Results of the irradiations of α-**10** and α-**40** by Powner *et al.*[323] showed similar minimal C1′ anomerisation, as reported by Sanchez and Orgel, but also identified some of the components formed (leading to loss of material), namely oxazolidinone formation, epimerisation to *arabino*-compounds and base loss (*Table 6.9*, entries 1-2). Additionally, Powner *et al.* proposed a mechanism to account for the formation of ribooxazolidinone *ribo*-**106**, the hydrolysis of the base moiety to form α-ribouridine (α-**56**), and the elimination of the bases leading to detection of the free bases cytosine **11** and uracil **29**[33] (*Scheme 6.12*).

Entry n°	Species irradiated	Time (h)	^{1}H NMR integration relative to other species integrated (%)								
			α-C	α-U	β-C	β-U	C[a]	U[a]	oxz[b]	α-ara	β-ara
1	α-**10**[*]	16	**51**	tr	6	tr	10	tr	27	0	0
2	α-**40***	15	**17**	3	16	tr	19	5	30	0	7
3	β-**115***	72	15	3	**32**	30	10	5	0	0	0
4	β-**38***	87	7	4	**26**	26	10	7	0	3	0
5	α-**38**[±]	72	**23**	7	22	14	17	14	0	tr	tr

*Table 6.9: Composition of species following the irradiation of nucleotides (20mM) α-ribocytidine (α-**10**), α-ribocytidine 5′-phosphate (α-**40**), β-2′-deoxycytidine (β-**115**), and α/β-ribocytidine 2′-phosphate (α/β-**38**). Using a 245nm mercury lamp for the specified amount of time, followed by relaxation (4 h), lyophilisation, refluxing in D_2O (16 h, 90°C), lyophilisation and analysis by ^{1}H NMR in D_2O. Recovery of starting material is shown in bold. *As reported by Powner et al.[323] [±]In the irradiation of α-**38**, ~30% ribose-2-phosphate was recovered alongside base loss, as reported by Powner and Sutherland[324]. [a]Resulting from nucleobase loss: cytosine (C) and uracil (U). [b]Oxazolidinone are assumed to be in the ribo-configuration.*

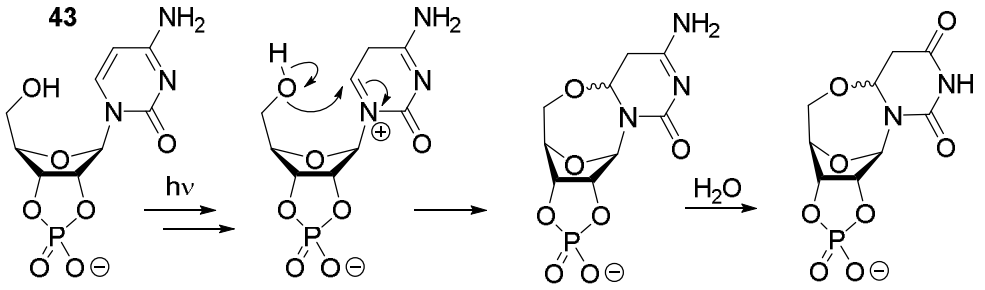


*Scheme 6.12: Mechanism for the photodestruction of α-ribocytidine (α-**10**), as proposed by Powner et al.*[33]*, leading to the formation of: ribooxazolidinone (ribo-**106**; red circle), base loss to cytosine (**11**), hydrolysis of the base to form α-ribouridine (α-**56**) and its base loss product uracil (**29**).*

This mechanism suggests that the C2′ alcohol is necessary for the formation of oxazolidinones (*Scheme 6.12*, red circle). If that is the case, then blocking that C2′-OH would prevent loss of material through this pathway. Powner *et al.* therefore studied the irradiation of α-ribocytidine 2′-phosphate (α-**38**)[324], as well as β-**38** and β-2′-deoxycytidine (β-**115**)[323]. Results showed that the main chemical pathways were those of hydrolysis of the cytidine base moiety to uridine, with recovery of starting material and significant nucleobase loss (*Table 6.9*, entries 3-5). However, excitingly, no oxazolidinones were observed, lending credence to

the proposed mechanism (*Scheme 6.12*). In the case of α-**38**, they also reported a significant increase in C1′ anomerisation (36% β-**38** and its uridine variant combined; *Table 6.9*, entry 5), suggesting that by blocking the degradation pathway through formation of oxazolidinones, a higher proportion of nucleotides could be retained.

In later studies, Powner *et al.*[33] found a way to avoid significant loss of material through base loss by irradiating β-ribocytidine-2′,3′-cyclic phosphate (β-**43**; 20mM, using a 245nm mercury lamp, pH 6.5) which lead to recovery of starting material (43% after 3.5 days) and partial hydrolysis of the base to yield β-ribouridine-2′,3′-cyclic phosphate (β-**57;** 42%), with only minor base loss (9%) observed during the experiment. The improved stability of cyclic phosphate β-**43** over 2'-phosphate β-**38** towards irradiation was explained by the strain experienced by the sugar moiety and caused by the ester linkage of the 2′,3′-cyclic phosphate[325], leading to β-**43** adopting a 'west' conformation[33]. This unusual conformation leads to a closer proximity of the C5′-OH and the nucleobase in β-anomers, leading to the possibility of photo-adducts, which is proposed to stabilise the base towards irradiation[33] (*Scheme 6.13*).



*Scheme 6.13: Mechanism proposed by Powner et al.[33] to account for the stability of β-ribocytidine-2',3'-cyclic phosphate (β-**43**) and β-ribouridine-2',3'-cyclic phosphate (β-**57**) during irradiation.*

The irradiation studies of **43** were performed for another reason than anomerisation: Powner *et al.* wanted to find a way to 'clean-up' a mixture of compounds obtained from previous steps, while selectively retaining the desired product β-**43**, and obtain the other canonical pyrimidine nucleotide β-**57** in the

process (by partial hydrolysis of the nucleobase). Indeed, when subjecting the α-anomer α-**43** to the same irradiations conditions (for 3.5 days), they find that it leads to photodestruction to (uncharacterised) non-nucleotide products, with traces of free nucleobases[33] (further confirming the theoretical mechanism of the base protection described in *Scheme 6.13*). Furthermore, irradiation of other products obtained from previous steps, including α-**40**, α-**10**, β-**51**, β-**113**, leads to significant loss of material through oxazolidinone formation or base loss and photodestruction[33]. Powner *et al*. therefore conclude that irradiation of their complex mixture of products from previous steps could lead to a simplification of the reaction mixture and effective amplification of the desired canonical nucleotides β-**43** and β-**57**.

All of these findings suggest that photoanomerisation of C1′ (α to β) could be possible if the C2′-OH is absent or blocked, but not by a cyclic 2′,3′ linkage, which seems to enhance photodestruction in α-anomers. These results encouraged us to take advantage of the overwhelming *ribo*-selectivity of our synthesis thus far leading to α-ribocytidine-5′-phosphate (α-**40**), in the hopes that we might find a prebiotically plausible way of selectively anomerising the C1′ to yield canonical pyrimidine ribonucleotides 5′-phosphate (β-**40**; and possibly the uridine variant β-**110**).

Repeating the irradiation experiments described above for α-**40**, at a range of concentrations, showed that α-**40** gets destroyed in a concentration dependent manner (*Table 6.10*)[xxv]. As seen by Powner *et al.*[323], oxazolidinone formation and nucleobase loss compete with the desired C1′ anomerisation, leading to low yields of β-anomers.

[xxv] The following results are part of unpublished work carried out by Dr. C. A. Fernandez in the Powner lab (*manuscript in preparation*).

Entry n°	Species irradiated	Conc. (mM)	^{1}H NMR integration relative to other species integrated (%)							
			α-C[a]	α-U[a]	β-C[a]	β-U[a]	C/U[b]	oxz[c]	α-ara	β-ara[d]
1	α-**40**	10	**48**	0	7	tr	15	26	tr	4
2	α-**40**	5	**14**	8	20	tr	16	21	tr	21
2	α-**40**	2.5	**17**	9	5	5	26	22	tr	16

*Table 6.10: Composition of species following the irradiation of α-ribocytidine 5′-phosphate (α-**40**; concentration as stated) using a 245nm mercury lamp for 16 h, followed by relaxation (1 h), lyophilisation, refluxing in D_2O (16 h, 90°C), lyophilisation and analysis by ^{1}H NMR in D_2O. Recovery of starting material is shown in bold. [a]Ribonucleotides. [b]Combined nucleobase loss: cytosine (C) and uracil (U). [c]Oxazolidinone are ribose oxazolidinone 5′-phosphate (ribo-**107**). [d]β-arabinonucleotide 5′-phosphates were observed as both cytidine and uracil.*

Temporarily blocking the C2′-OH without compromising the 5′-phosphate would possibly afford the same anomerisation amplification observed in the 2′-phosphate α-**38** irradiations. The protecting group used would need to be prebiotically available, small and easily removable by prebiotic means. Protecting group strategies, such as acetylations, are often used in non-prebiotic nucleotide chemistry[326], and the simplicity of acyl-transfer reactions made it a good candidate for a prebiotic protection of α-**40**. There is precedent for the use of acetylation in prebiotic chemistry[327], such as the strategy developed by the Sutherland group and used to regioselectively direct 3′-5′ oligomerisation by blocking the 2′-OH (thereby blocking any 2′,3′ cyclic phosphate formation, *vide supra*)[319]. Proof-of-concept experiments *via* non-prebiotic, "standard" acetylation of α-**40** (in pyridine with acetic anhydride), afforded near-quantitative formation of tri-acetylated α-**116**[322]. Irradiation of α-**116** did not lead to any noticeable C1′ anomerisation (*Table 6.11*, entry 1), instead going through destructive pathways: nucleobase loss and hydrolysis of the base moiety to form the uridine variant α-**110**.

Entry n°	Species irradiated	Conc. (mM)	^{1}H NMR integration relative to other species integrated (%)						
			α-C[a]	α-U[a]	β-C[a]	β-U[a]	C/U[b]	oxz[c]	β-ara[d]
1	α-**116**	10	**24**	51	tr	tr	19	6	0
2	α-**117**	10	**92**	tr	2	0	6	tr	0
3	α-**117**	5	**59**	15	14	0	7	5	0
4	α-**117**	2.5	**41**	19	18	4	11	7	0

*Table 6.11: Composition of species following the irradiation of nucleotides (concentration as stated): 2',3'-O-diacetyl-α-ribocytidine 5'-phosphate (α-**117**) and triacetylated α-**116**. Using a 245nm mercury lamp for 16 h, followed by relaxation (1 h), lyophilisation, refluxing in D_2O (16 h, 90°C), lyophilisation and analysis by ^{1}H NMR in D_2O. Recovery of starting material is shown in bold. [a]Ribonucleotides. [b]Combined nucleobase loss: cytosine (C) and uracil (U). [c]Oxazolidinone are ribose oxazolidinone 5'-phosphate (ribo-**107**).*

Interestingly, aqueous acetylation using prebiotically plausible reaction methods[319,328] showed markedly different results. Incubation of α-**40** (100mM in water, near neutral pH) in the presence of either *N*-acetylimidazole (10 eq), or thioacetic acid (10 eq) and ferricyanide (8 eq), showed selective formation of 2',3'-*O*-diacetyl-α-ribocytidine-5'-phosphate (α-**117**; in 91% and 80% yield respectively), with no sign of *N*-acetylation of the nucleobase moiety. Studying the irradiation of the prebiotically acetylated compound α-**117** at a range of concentrations (in similar conditions as previous irradiations[xxvi]) showed promising results as up to 22% of the canonical β-**40** and β-**110** were obtained (*Table 6.11*, entry 4). Furthermore, only little oxazolidinone formation and no C2′ epimerisation (leading to arabinonucleotide), was observed, thereby minimising product loss and conserving the canonical *ribo*- configuration (*Figure 6.18*, *b.*).

[xxvi] α-**117** (2.5-10mM) at pH 6.5 in water, sparged with Ar or N_2, was irradiated with a UV lamp (254nm principle emission) for a set amount of time. The irradiation was followed by lyophilisation, re-suspension in water and heating for 16 h at 90°C, followed by de-acetylation (incubation with NH_4OH for 16 h at RT) and lyophilisation. The lyophilite was then analysed by NMR in D_2O.

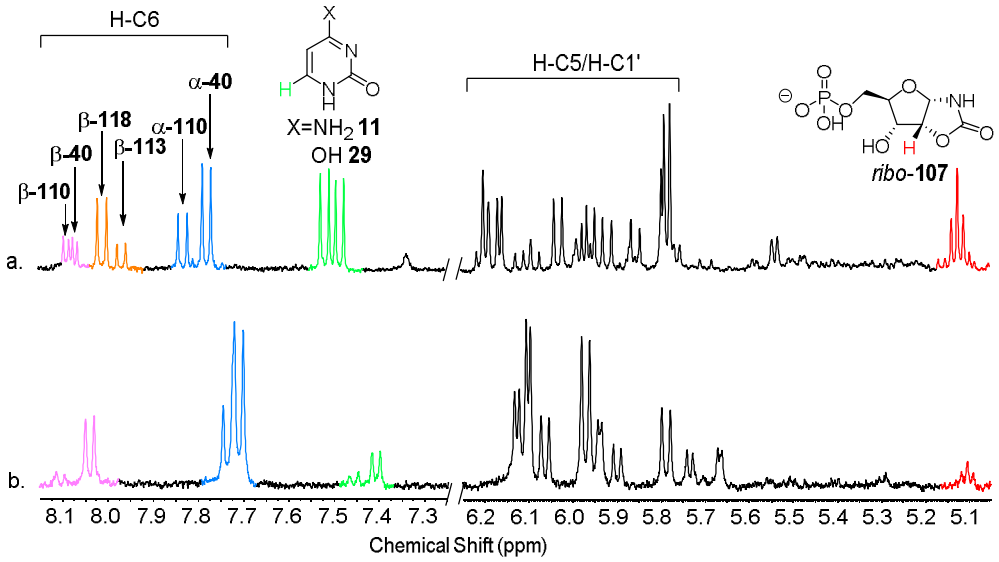


*Figure 6.18: ^{1}H NMR spectra (400 MHz, D_2O, 8.15-5.00 ppm) of the irradiation of nucleotides (2.5mM; 254nm principle emission) for 16 h followed by heating at 90°C for 16 h, lyophilisation and resuspension in D_2O. a. Irradiation of α-ribocytidine 5'-phosphate (α-**40**). b. Irradiation of 2',3'-O-diacetyl-α-ribocytidine 5'-phosphate (α-**117**). Reaction was analysed after de-acetylation (incubation with 28 % NH_4OH in water, for 16 h at RT) and lyophilisation. The lyophilite was then analysed by ^{1}H NMR in D_2O. Regions of interest are annotated to show evidence (or lack thereof) of C1′ anomerisation (β-**40**/β-**110**), C2′ anomerisation (β-**113**/β-**118**), nucleobase loss (cytosine (**11**)/uracil (**29**)), and oxazolidinone 5'-phosphate (**107**) formation.*

These results show that bis-acetylated compound α-**117** undergoes C1′ anomerisation to form canonical nucleotides 5'-phosphate β-**40**/β-**110** (following NH_4OH deprotection). In contrast, non-prebiotically tri-acetylated compound α-**116** undergoes significant nucleobase photohydrolysis to yield the uridine variant, with only trace amounts of C1′ anomerisation. This suggests that while 2′- and 3′-OH protection does increase the anomerisation potential of 5′-phosphorylated α-ribocytidine compounds, acetylation of the nucleobase moiety is detrimental. Furthermore, α-**117** irradiations show potential regarding the prebiotic synthesis of β-**40**/β-**110**, as detrimental C2′ epimerisation was completely avoided and oxazolidinone synthesis was greatly minimised, leading to augmented C1′ anomerisation and recovery of the starting material. The yields obtained, although moderate, are of the same order as the previous highest

reported yield in the anomerisation of α-ribocytidine 2′-phosphate (α-**38**)[324]. However, in contrast with α-**38**, bis-acetylated α-**117** is easily synthesised (and deprotected in prebiotic conditions). Indeed, the prebiotic protecting group strategy of acetylation allows for a prebiotic temporary protection of the problematic 2′-OH (along with the 3′-OH) while keeping the advantageous 5′-phosphate group intact. These irradiation experiments complement our selective prebiotic synthesis of 5′-phosphate riboanhydrocytidine (*ribo*-**112**), thus demonstrating a comprehensive prebiotic synthesis of canonical 5′-phosphorylated pyrimidine ribonucleotides, which can be used for efficient oligomerisation.

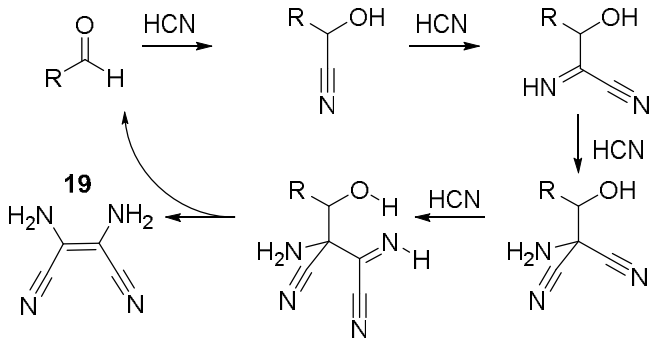
7. Purine synthesis

7.1. Towards a prebiotic synthesis of purine nucleotides

A prebiotically plausible pathway to purine nucleotides remains one of the most significant problems associated with understanding the origins of the RNA chemical subsystem of life[274]. Indeed, while a plausible pathway to pyrimidine ribonucleotides has been established through decades of work[227,270,310,329], culminating in the development of a 5 step prebiotic synthesis by Powner *et al.*[33] (as discussed extensively in previous sections), no comprehensive synthesis of the purine counterpart has yet been achieved[274].

Successful, albeit low yielding, prebiotic synthesis of the purine bases has been achieved, with the view of later performing a glycosidation reaction with ribose (**9**) to form ribonucleotides (a point to be returned to later). Synthesis of adenine (**18**) was demonstrated by Oró in 1960 by oligomerisation of HCN in the presence of ammonia. Oró observed a 0.5% yield in the reaction of 2.1M HCN in 3M $NH_4OH_{(aq)}$, after several days at 70°C (followed by hydrolysis with 6M HCl), the rest of the material was described as insoluble polymer[191,192]. These results were later confirmed by Lowe *et al.*[124] who reported the formation of **18** in similar conditions as well as the production of a complex mixture of organic compounds relevant to prebiotic synthesis (including amino acids and other purine nucleobases). The work was however heavily criticised for the conditions used: high concentrations of cyanide and ammonia, and harsh and incompatible conditions (heating in strong base and subsequent hydrolysis in cool acid)[166]. Furthermore, the presence of large quantities of ammonia on the early earth has been debated[195,196]. As such, other syntheses by-passing the need for concentrated ammonia have been carried out, including those by Schwartz *et al.*[197] which was carried out in ice. Eutectic phases of ice allow low concentrations of HCN

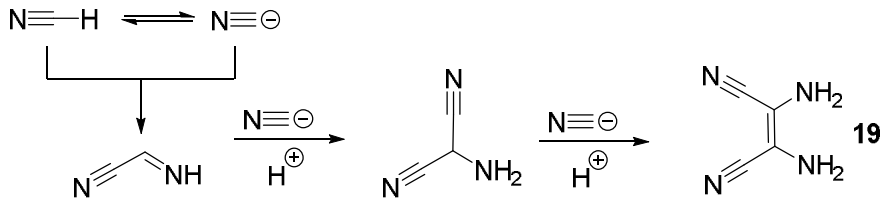
(10mM) to be used[xxvii], argued to be more prebiotically plausible than the high concentrations used in previous studies. They reported low yields of adenine (**18**) produced (0.004%). Interestingly, Schwartz observed an increased (0.02%) yield of **18** by addition of glycolonitrile (**60**)[197]. The increased rate formation of HCN oligomerisation in the presence of aldehydes is proposed to occur due to the formation of cyanohydrins of the aldehyde (a near quantitative, extremely efficient reaction)[198]. The formation of a cyanohydrin facilitates the oligomerisation of HCN by providing a more electrophilic nitrile to initiate the oligomerisation process (*Scheme 7.1*), thereby avoiding the slow addition of cyanide to HCN (a reaction that is described as the rate limiting step)[330]. The essential retro-aldolisation step then regenerates the aldehyde catalyst while liberating the product of HCN oligomerisation, HCN tetramer diaminomaleonitrile (DAMN; **19**)[330].



*Scheme 7.1: Mechanism for the aldehyde-catalysed oligomerisation of HCN, forming tetramer diaminomaleonitrile (**19**), as observed by Schwartz et al.[198], adapted from Eschenmoser*[330].

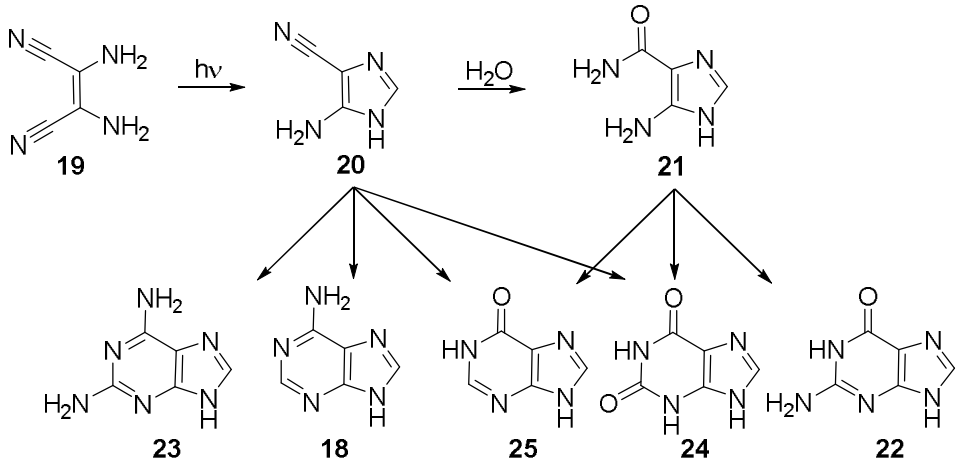
Ferris and Orgel demonstrated a different pathway to purines, *via* photo-isomerisation of dilute HCN (0.3mM) to efficiently form purine precursor, HCN tetramer **19** (highest conversion rate occurred at the pK_a of HCN, at 25°C: pH 9.2)[199] (*Scheme 7.2*).

[xxvii] The low concentrations of HCN would be increased by freezing out the competing solvent, water.



*Scheme 7.2: Production of diaminomaleonitrile (**19**) from HCN, adapted from Orgel and Lohrmann*[169].

They also report high yielding conversion of **19** to 4-amino-imidazole-5-carbonitrile (AICN; **20**) (77-82% when 0.1mM **19** is irradiated at 25°C) while suggesting a plausible mechanism for the production of purines from this precursor[200] (*Scheme 7.3*). Sanchez *et al*. later demonstrated that purines can be obtained by heating **20** (100mM, 100°C, 4 days) in the presence of ammonium cyanide (1M NH_4CN, pH 9-10), leading to **18** (11%) or hypoxanthine (**25**; 3%). The latter is formed when starting with the hydrolysis product of **20**: 4-amino-imidazole-5-carboxamide (AICA; **21**). Additionally, they found that reactions of **21** (10mM) with cyanogen (**5**; 10 eq) in the presence of ammonia (2-10 eq) led to high yield (41-43%) of guanine (**22**) (or 31% diaminopurine (**23**) when starting with **20**) when heating the mixture for a day at 100°C. A similar reaction with potassium cyanate (0.5M) also led to **22** production, or xanthine (**24**; when starting with **20**; *Scheme 7.3*)[201]. However, the low solubility of the purine base raises problems concerning their accumulation in water.



*Scheme 7.3: Reported prebiotic purine synthesis via HCN tetramer DAMN **19** leading to purine precursors AICN **20** and AICA **21**, and purines: adenine (**18**), hypoxanthine (**25**), xanthine (**24**), diaminopurine (**23**) and guanine (**22**). As reported by Sanchez, Ferris and Orgel*[201].

In order to form purine nucleosides, the pre-formed purine bases would need to be regiospecifically attached to pre-formed D-ribose (D-**9**). Racemic **9** has been shown to form through the formose reaction, discovered by Butlerov in 1861[170]. It has been proposed that formaldehyde (**2**) self-condenses initially to form glycolaldehyde (**3**), catalysed by calcium hydroxide, and then subsequently, reacts by aldol condensation to form more complex sugars[171], including minor amount of **9** (<1%)[180]. However, this reaction also forms many other sugars and their degradation products, as can be seen in the gas chromatography studies acquired by Decker *et al.*[180] (*Figure 7.1*), which would all compete in the theoretical formation of RNA-type polymers.

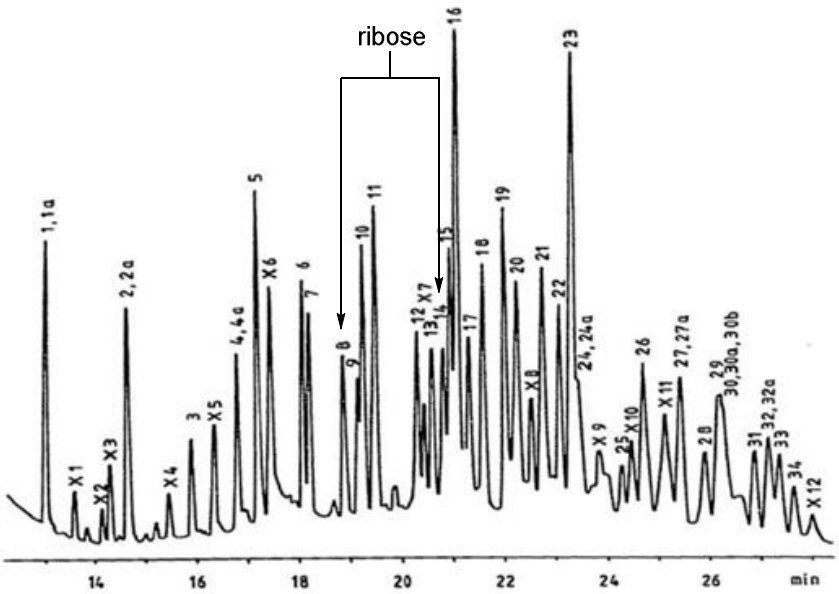


Figure 7.1: Gas chromatogram of products of n-butoxime derivatives of the formose reaction, as shown by Decker et al.[180]

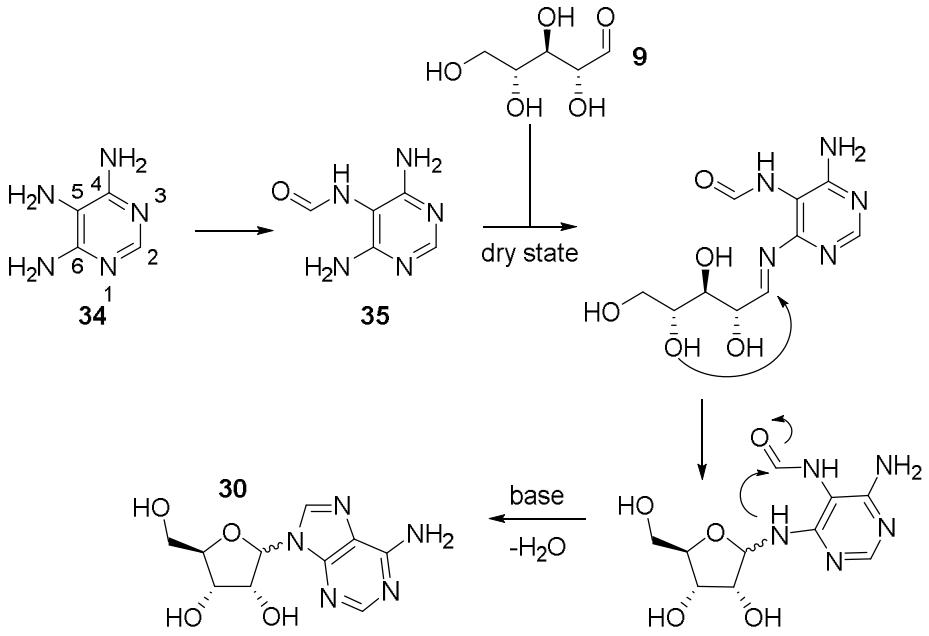
Additionally there are reports that with carefully distilled formaldehyde (**2**), no formose reaction takes place, which suggests that the formose reaction occurs due to trace impurities of "biologically derived sugars" in **2**[179]. A wide variety of additives have been sought to improve the selectivity of formation of ribose (**9**)[63]. Three notable examples include (as mentioned in *Section 1.2.2*): ribose (**9**) stabilisation by borate[184], zinc-proline catalysed aldol reaction of glycolaldehyde (**3**) and glyceraldehyde (**12**)[185,186,331], and glycolaldehyde-phosphates crossed aldomerisation with formaldehyde (**2**)[332]. Ricardo *et al.*'s work suggests increased pentose formation in the presence of borate, however the stereoselectivity is not reported[184]; Kofoed *et al.*'s work on zinc-proline catalysed reactions shows ribose (**9**) as being 20% of the final reaction mixture[185,186,331], while Pitsch *et al.*'s work on mineral induced formation of sugars shows 31% yield of ribopyranosyl (after enzymatic dephosphorylation)[332]. However, it transpired that before dephosphorylation ribose was found with an uncanonical phosphorylation pattern: ribose-2,4-diphosphate **17**, and no mechanism to date suggests that a conversion from 2,4-ribopyranose diester to a canonical 3,5-ribofuranose diester could occur.

Even though some prebiotic syntheses of ribose (**9**) have been achieved, the problem remains that it is unstable in aqueous solution with a half-life of only 300 days at 25°C and neutral pH[188], thus making it essentially unavailable for prebiotic use, compelling some to say that the first genetic material could not have been generated from **9**[188]. Furthermore, it has been noted by Shapiro that "nitrogenous substances (needed for prebiotic base synthesis) would interfere with the formose reaction by reacting with **2**, the intermediates, and sugar products in undesirable ways."[190], putting into question the prebiotic plausibility of the selective synthesis of pure **9**.

The remaining problem in prebiotic synthesis of purine ribonucleotides (assuming a route to pure nucleobases and pure **9**) is the selective glycosidation of nucleobases. Early glycosidation experiments, largely pioneered by Orgel and co-workers, found that adenosine (**30**) and inosine (**31**) nucleosides could be assembled from pure D-ribose (D-**9**) and the corresponding purine nucleobases albeit in low yields: after a screen of conditions, only a maximum of 9% yield can be made for β-**31**, and only 4% yield for β-**30**[208]. They also reported an 8% yield of β-guanosine (**32**)[208]. The biggest problem arising from these syntheses is the fact that the small amounts of canonical nucleotides were contaminated with all other possible anomers and regioisomers: *N*7-, *N*9-, NH_2-C6- substituted glycosides in both the α- and β-anomers and in both pyranosyl and furanosyl ring conformations. The lack of selectivity of these glycosidation methods would lead to significant problems during oligomerisation and the expression of function of oligonucleotides, as anomers and isomers would inhibit synthesis and function of a biopolymer.

Recent improvements have been made on these glycosidations, by using a precursor to purine nucleobases in the glycosidation step: triaminopyrimidines (4,5,6-triaminopyrimidine (**34**) or variants)[213]. The marked advantage of using these bases is their increased solubility in water. Becker *et al.*[213] found that by using **34**, they could limit the stereochemistry of the glycosidation to the desired nitrogen *N*9. Indeed, they found that the exocyclic nitrogen *N*5 was more

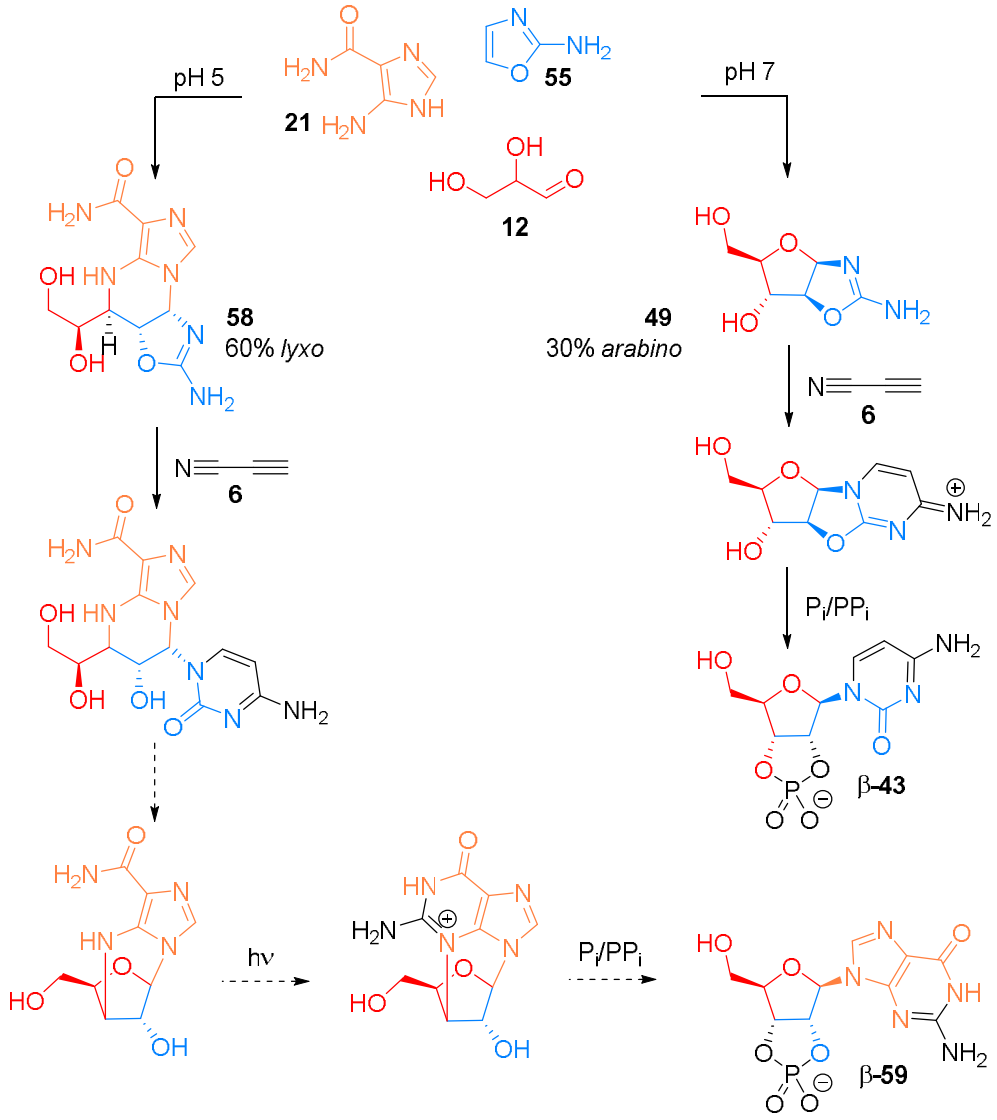
nucleophilic by virtue of its greater sp^3 character (the other two exocyclic nitrogens, *N*4 and *N*6 being "protected" by the in-ring nitrogen's protonation at slightly acid pH), leading it to be selectively activated by formic acid or formamide (**36**) to yield *N*5 formylated triaminopyrimidine **35** (*Scheme 7.4*). Reacting **35** with pure ribose (**9**) (in the dry state, followed by stirring in slightly basic conditions) yielded selective *N*9 glycosidation with the remaining exocyclic nitrogen (*N*4 or *N*6[xxviii] become *N*9 once attached to the sugar moiety) being free to react to yield adenosine (α/β-**30**; in both the furanosyl and pyranosyl form). The highest yield of the canonical furanosyl β-**30** was 20% (when following the dry-state procedure with refluxing with sodium borate: 0.125M at 100°C for 48 h). While this procedure significantly improves the yield of β-**30**, it in no way solves the aforementioned problem of the source of pure **9** or the lack of regioselectivity observed from the sugar moiety, as the numerous non-canonical pyranosyl and α-anomers formed can still interfere with the synthesis and function of oligomerised RNA.



*Scheme 7.4: Pathway to adenosine (**30**) via 4,5,6-triaminopyrimidine (**34**) as reported by Becker et al.*[213]

[xxviii] *N*4 and *N*6 are identical due to C2-C5 plane of symmetry in **35**.

Another work of note concerning purine nucleotide synthesis is a multicomponent synthesis with promising results concerning the one-pot concomitant, yet chemically divergent, synthesis of pyrimidine precursors and potential purine precursors carried out by Powner *et al.*[274] (*Scheme 7.5*). This work seeks to avoid the problematic glycosidation step while increasing the complexity of compounds in a prebiotic and selective manner. The purine synthesis they started developing utilises similar pathways and starting materials to the ones described in the novel prebiotic synthesis of pyrimidines[33] (described throughout the previous sections, and summarised in *Scheme 7.5*). One compound central to the building of the sugar moiety of the pyrimidine nucleotides is 2-aminooxazole (**55**; formed from the reaction of glycolaldehyde (**3**) and cyanamide (**4**), catalysed by P_i), which Powner *et al.* propose could also be used in the reaction of purine precursors AICA **21** or AICN **20** with aldehydes, leading to a "tethered aminoimidazole nucleotide intermediate"[274] such as **58** (*Scheme 7.5*). The findings they report show that a multicomponent reaction of **55**, **21** and glyceraldehyde (**12**) yielded the desired 3-component product *lyxo*-**58** (60% by ^{1}H NMR analysis; *Scheme 7.5*) at lower pH (pD 5), while at higher pH a divergent path yielded the product of the bimolecular reaction of **55** and **12**, leading to pentose aminooxazolines (**49**; with a high *ribo*- and *arabino*- preference, 44% and 30% respectively). The high stereoselectivity observed in both parallel pathways is of importance as they can lead to the canonical stereochemistry in each end product: β-ribocytidine-2′,3′-cyclic phosphate (β-**43**; and the uridine variant β-**57**) from *arabino*-**49**, and the proposed development to β-riboadenosine-2′,3′-cyclic phosphate (β-**59**) from *lyxo*-**58** (*Scheme 7.5*).

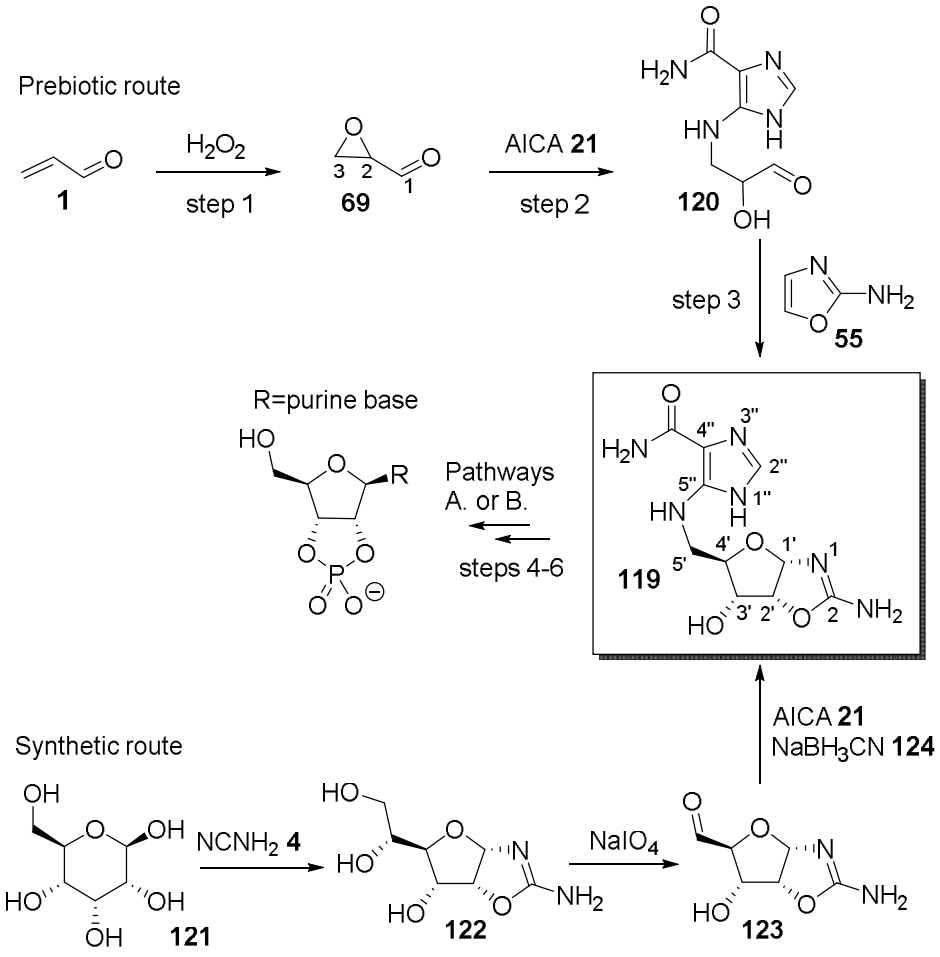


*Scheme 7.5: Divergent pathways leading to pyrimidine precursors (2-component product, pentose aminooxazolines (**49**)) and potential purine precursors (3-component product **58**) in the pH dependent reaction of glyceraldehyde (**12**), 2-aminooxazole (**55**) and AICA **21**, as proposed by Powner et al[274]. Observed chemistry in solid lines, hypothetical chemistry in dotted lines.*

This novel 3'-purine-tethering strategy proposed by Powner *et al.*[33,274,333] would by-pass the need for glycosidation while allowing for a pH-depend distribution of products leading to canonical purine and pyrimidine nucleotides[xxix]. We propose to investigate the formation of canonical purines *via* several different pathways,

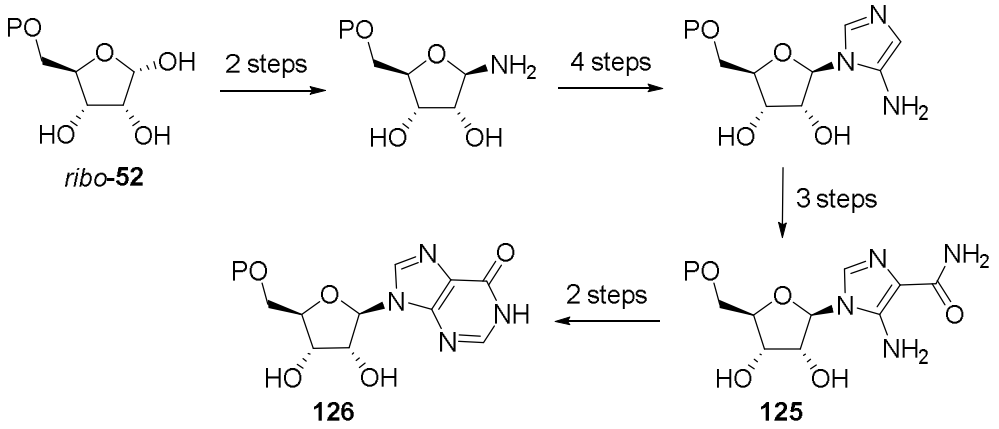
[xxix] This strategy still forms a part of the ongoing investigation into the prebiotic synthesis of ribonucleotides in the Powner Lab.

centred on the theme of purine precursor tethering at the 5'-carbon atom of sugar synthons (*Scheme 7.6*). This chemistry will be based on the reactions of simple C_2, C_3 units, and the combination of the previously distinct nitrogenous and oxygenous chemistry. We wish to establish a prebiotic chemical method to stereo- and regio-specifically introduce a purine (or precursor) aglycone at the anomeric centre of purine nucleotides. Our proposed routes would bypass the need for unstable free-sugars and nucleobases, and overcome the difficulties associated with regioselective glycosidation (*vide supra*).



*Scheme 7.6: Proposed synthetic and prebiotic pathways to synthesise key intermediate 5'-AICA riboaminooxazoline (**119**), and its subsequent elaboration into purine ribonucleotides.*

The C_5 sugar moiety of ribonucleotides can be accessed through C_2+C_3 units or C_2+C_2+C_1. 2-Aminooxazole (**55**; prebiotically synthesised from C_2 glycolaldehyde (**3**) and cyanamide (**4**)) can be used as a C_2 unit to build the sugar ring through masked-aldol reaction with glyceraldehyde (**12**; or variant). **55** has been used as a key intermediate in both the reported pyrimidine synthesis and reported 3'-purine-tethering strategies proposed by Powner *et al.*[33,274,333] (*Scheme 7.5*, *vide supra*), and we propose to include it in our novel 5′-purine-tethering route for the same purpose. Our proposed prebiotic route to purine nucleotides beginnings with oxidation of prebiotically plausible acrolein (**1**) to glycidaldehyde (**69**) (as described in *Section 2*; *Scheme 7.6,* step 1)[284]. We seek to demonstrate that aglycone precursors of purine nucleobases such as AICA **21** can be tethered to glyceraldehyde-synthon **69** at C3 to form glyceraldehyde-3-AICA (**120**; step 2). **21** is an important intermediate *en route* to purine bases from HCN (*vide supra*) and we propose that the densely functionalised structure of **21** can be used to regiospecifically direct purine glycosidation. The product of this triose-purine precursor tethering, **120**, could then undergo a regiospecific masked aldol reaction with **55** to form key intermediate 5′-AICA riboaminooxazoline (**119**), thereby building the sugar moiety of the nucleotide (step 3). We anticipate sugar synthesis will occur with complete furanosyl selectivity based upon the previously reported reactions of **55** with **3**, **12**, glyceraldehyde-3-phosphate (**68**) and **69**[272,274,283,333] (reactions of **55** and **69** are described in *Section 4*). **119** will be a key compound in this study as a 5'-AICA analogue of known nucleotide precursor riboaminooxazoline (*ribo*-**49**), which, additionally, bears strikingly close relationship to AICA-riboside **125**, a metabolite in the *de novo* purine biosynthesis (*Scheme 7.7*).

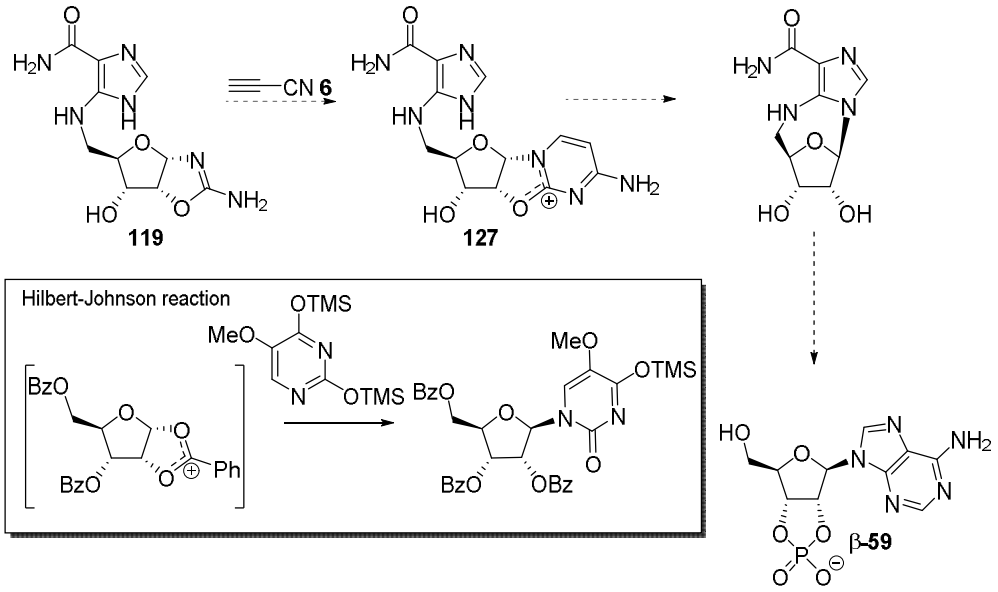


Scheme 7.7: De novo synthesis of purine ribonucleotide inosine mono-phosphate (IMP) ***126*** *via enzyme mediated steps starting with ribose 5-phosphate (ribo-**52**) and featuring AICA-riboside (**125**).*

As mentioned previously, glycosidation of free-bases by ribose (**9**) has been previously reported, but is low-yielding. These conditions inevitably lead to other non-natural nucleosides as no preference for the canonical nucleotide anomer or regiochemistry has been observed, reducing further the yield of the desired natural nucleosides. We seek to overcome the problems associated with non-selective glycosidation to effect a selective β-glycosidation of the *N*9-nitrogen atom of purines (n.b. numbered *N*1″ when it is still in its tethered AICA stage; *Scheme 7.8*).

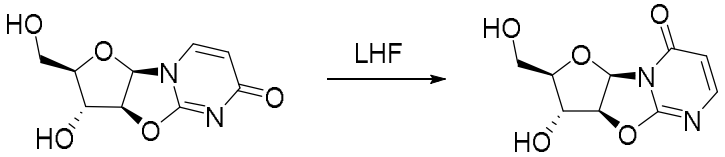
For the proposed prebiotic conversion of 5′-AICA riboaminooxazoline (**119**) to canonical purine nucleotides, three further steps need to be achieved: activation of **119** to glycone exchange, 5'-displacement, and purine elaboration (*Scheme 7.6*). We envisage purine elaboration as a flexible step in this synthesis, occurring either before or after 5′-displacement. We propose two specific pathways for initial investigation (A. and B., *vide infra*), diverging after **119** formation and both involving purine elaboration after 5′-displacement.

Route A. Nucleophilic displacement of anhydropyrimidine, which we believe is (partially) analogous to the delocalised oxonium intermediate exploited in neighbouring group directed Hilbert-Johnson glycosidation (*Scheme 7.8*).



*Scheme 7.8: Proposed prebiotic stereochemical control of purine base insertion via a 5'-tethered Hilbert-Johnson analogue 5'-AICA riboanhydrocytidine (**127**). In box: Standard Hilbert-Johnson reaction.*

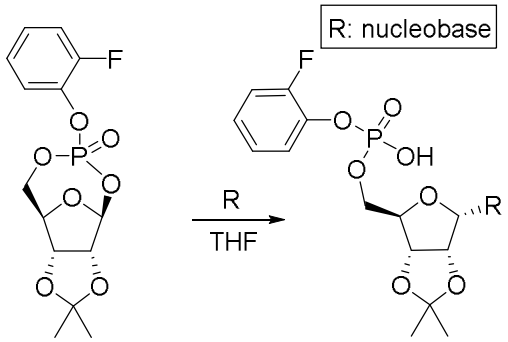
The positively charged anhydronucleoside **127** is anticipated to have a similar reactivity (albeit much less so) to the proposed positively charged intermediates resulting from neighbouring group participation during Hilbert-Johnson glycosidations. Partial literature precedent for the displacement of pyrimidine *N*1 can be found in the isomerisation of 2,2'-anhydrouridine to 2,2'-anhydro-pseudo-uracil in strongly acidic liquid hydrogen fluoride[334] *(Scheme 7.9)*.



Scheme 7.9: Anhydrouridine N1 displacement, followed by rearrangement to N3 anhydro-pseudo-uracil in liquid hydrogen fluoride (LHF), adapted from Polazzi et al[334].

We hypothesise that pyrimidine-activation and chemical tether of AICA **21** will lead to stereochemical inversion at the anomeric position, resulting in a

β-glycosidation, as is observed in all naturally occurring RNA. Literature precedent for participation of a nucleophilic 5'-moeity is found in the work of Mukaiyama *et al.* who demonstrate that 5'-phosphate triesters are observed in Hilbert-Johnson-type stereoselective glycosidation reactions[335] (*Scheme 7.10*).



Scheme 7.10: Hilbert-Johnson type stereoselective glycosidation of nucleoside bases in THF, adapted from Mukaiyama et al.[335]

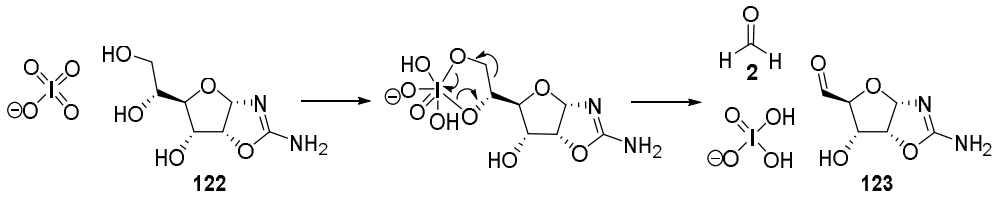
Route B. Photochemical pyrimidine activation. We hypothesise that 5'-carbon-tethered ACIA/N nucleotide precursor will control regio-selective glycosidation. As discussed in *Section 6*, the photochemical irradiation of pyrimidine nucleotide liberates the free-nucleobase and, if 2'-phosphorylated, the corresponding sugar is also isolated[323]. Therefore we seek to investigate selective phosphorylation of the vicinal diol, to give 2',3'-cyclic phosphate[224] (as discussed in *Section 6*) and study photochemical pyrimidine displacement in the presence of a tethered purine moiety.

Each pathway then ends similarly with elaboration of the purine precursor – through modification of published purine annulation protocol. For example Sanchez and coworkers have reported the synthesis of adenine (**18**) and diaminopurine (**23**) from AICN **20** and guanine (**22**) from AICA **21** in 11%, 31% and 43% yield, respectively[201], to give us access to both the purine nucleotides. The final step of our proposed route will be intermolecular displacement of the *N*3-purine tether. These displacement also have literature precedent as they have previously been used by Mizuno *et al.* to synthesis 3'-5'-ligated oligonucleotides[336].

7.2. Synthetic studies of 5′-substituted aminooxazolines

7.2.1. Synthesis of 5′-oxoriboaminooxazoline

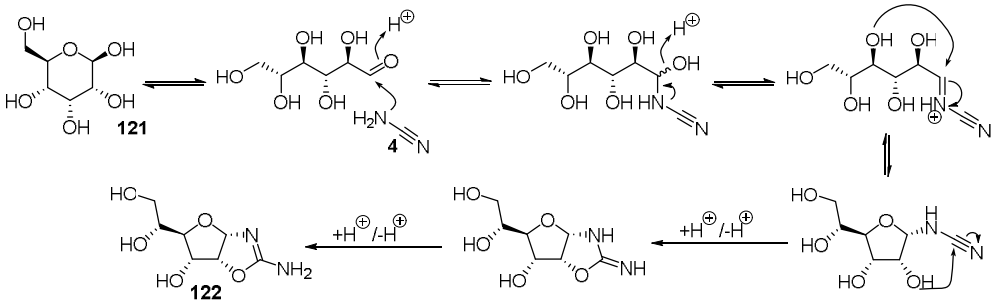
As described earlier, our first goal was to build key intermediate 5′-AICA riboaminooxazoline (**119**) efficiently *via* a scalable chemical pathway, both as a chemical standard for our prebiotic synthesis (*vide infra*) and as a means to rapidly investigate the forward pathway from **119**. Our initial retrosynthetic analysis of **119** suggested that reductive amination of aldehyde 5′-oxoriboaminooxazoline (**123**) would provide rapid scalable access to **119** (*Scheme 7.6*, *vide supra*). The first route envisaged to synthesise target aldehyde **123** efficiently was periodate cleavage of novel aminooxazoline alloaminooxazoline (**122**; *Scheme 7.11*).



*Scheme 7.11: Mechanism of sodium periodate cleavage of alloaminooxazoline (**122**) to yield 5′-oxoriboaminooxazoline (**123**) and formaldehyde (**2**).*

We found that D-alloaminooxazoline (**122**) was readily synthesised upon treatment of commercially available D-allose (**121**) with cyanamide (**4**) in the presence of catalytic (3.5% aq.) ammonia in a modification of the procedure reported by Sanchez and Orgel for the synthesis of riboaminooxazoline (*ribo*-**49**)[270] (*Scheme 7.12*). In time course reactions of **121** (130mM) at varying temperatures (RT to 80°C), with two and four equivalents of **4**, **122** was formed in good yield (87% crude **122**, as confirmed by ^{1}H NMR analysis of the above reaction with 4 eq **4**, heated at 80°C for 1 h). The rapidity and selectivity of the reaction at 80°C with four equivalents of **4** was chosen as a basis for scale-up to

gram scale production of **122**. At gram scale, product **122** crystallised directly upon cooling to 4ºC to afford 645 mg of white powder (58%, *Figure 7.1, a.*). Interestingly, an in depth NMR analysis of the product demonstrated that the complete furanosyl selectivity was shared with that observed in the synthesis of ribose and arabinose aminooxazolines (*ribo*-**49** and *arabino*-**49**; but not lyxose aminooxazoline (*lyxo*-**49**) – which is observed to be formed as a mixture of pyranosyl and furanosyl isomers, that favours the pyranosyl product at equilibrium, 5:1 ratio[272]). This furanosyl selectivity is essential to our synthetic plans as we ultimately want to produce a purine nucleotide with natural configuration.



*Scheme 7.12: Proposed mechanism for novel aminooxazoline formation: D-alloaminooxazoline (**122**), from reaction of D-allose (**121**) with cyanamide* (**4**).

Once D-alloaminooxazoline (**122**) was purified by crystallisation, the next step involved oxidative cleavage of the 5',6'-diol with sodium periodate to form 5′-oxo-riboaminooxazoline (**123**; *Scheme 7.11*, *vide supra*), following a modification of Paquette's oxidative cleavage of 4,6-*O*-benzylidene-D-glucose[337], optimised for **122**. We encountered a number of purification difficulties due to the polarity of the compound, and various means were tested including crystallisation with various solvents and co-solvents, sodium bisulphite precipitation, and finally flash column chromatography (FCC). FCC proved most suitable to purify **123** on a small scale and after optimisation of elution condition (5:93:2 to 30:68:2, MeOH:EtOAc:Et_3N) we isolated **123** in 36% yield. ^{1}H NMR analysis of purified **123** in D_2O (*Figure 7.1, b.*) showed clear loss of the C6′ protons, consistent with successful periodate cleavage. Furthermore, no aldehyde signal was observed for **123** suggesting it behaves much like glycolaldehyde (**3**) and glyceraldehyde (**12**)

(and ribose (**9**) when in its open-chain form[209]) in that it preferentially sits as the hydrated form in water (*Figure 7.1*, *b.*) instead of the aldehydic form. This hydration is however reversible and as such does not change the reactivity of the aldehyde significantly. The presence of the aldehyde functional group was nonetheless verified by a positive TLC dip tests with 2,4-dinitrophenylhydrazine (reacting with carbonyls but not alcohols). The test was also performed on positive and negative controls (glycolaldehyde (**3**) and riboaminooxazoline (*ribo*-**49**) respectively).

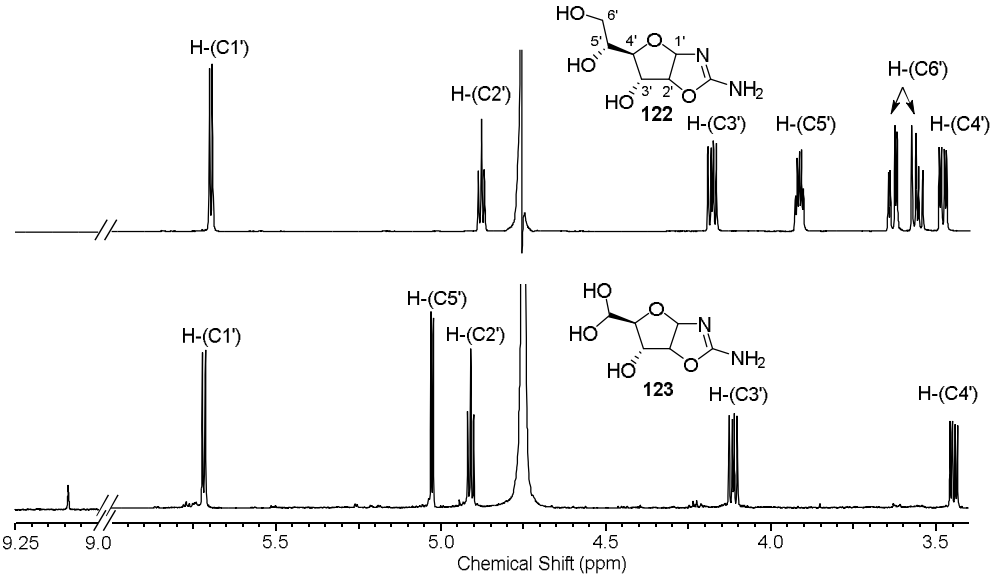
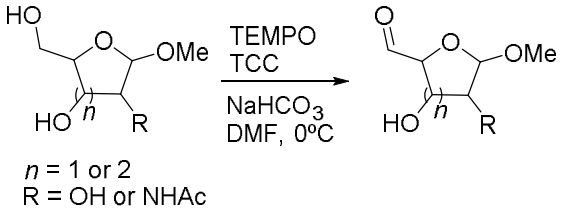


*Figure 7.1: Comparison of ^1H NMR spectra (600 MHz, D_2O, 9.25-3.40 ppm) showing cleavage of the C6' and subsequent loss of the H_2-C6' peaks in 5'-oxo-riboaminooxazoline (**123**). a. D-alloaminooxazoline (**122**), purified by crystallisation; b. Oxidative cleavage of **122** (0.15M) with sodium metaperiodate (1.5 eq) in water at RT, 2 h and subsequently purified by small scale column chromatography to yield **123**. Reactions and NMR were performed in water and D_2O respectively, leading to the aldehyde reversibly converting to the hydrate species, and therefore a loss of the downfield aldehyde peak in favour of a hydrate peak shown here around 5.0 ppm.*

Due to the limited availability of D-allose (**121**; £279.2/g, Sigma-Aldrich), an investigation of selective oxidation of riboaminooxazoline (*ribo*-**49**; 5mM) with TEMPO/TCC (cat./0.7 eq) in DMF, to access aldehyde **123** from D-ribose (**9**; £2.7/g Sigma-Aldrich) was carried out. Angelin *et al*. showed selective oxidation

of primary alcohols in unprotected glycosides can take place through use of mild conditions (dilute - 5mM starting material - at 0°C, with a catalytic amount of TEMPO)[338] (*Scheme 7.13*).



Scheme 7.13: TEMPO-catalysed oxidation of unprotected glycosides, selective for primary alcohols, adapted from Angelin et al[338].

Following the exact procedure (*vide supra*) with *ribo*-**49** showed promise as ^{1}H NMR analysis showed the presence of aldehyde **123**, albeit in lower yield than expected: 20% conversion (by ^{1}H NMR integration) compared to the 65% yield reported by Angelin *et al.* for their product[338]. However, 76% of species present in solution after 7 h was the starting material *ribo*-**49**, showing the reaction did not go to completion and perhaps slight modification of the reaction conditions will produce better yields (*Figure 7.2*)[xxx].

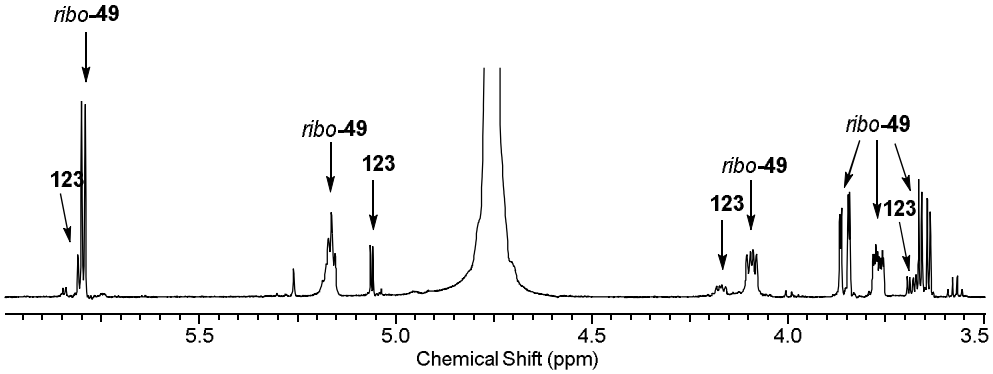
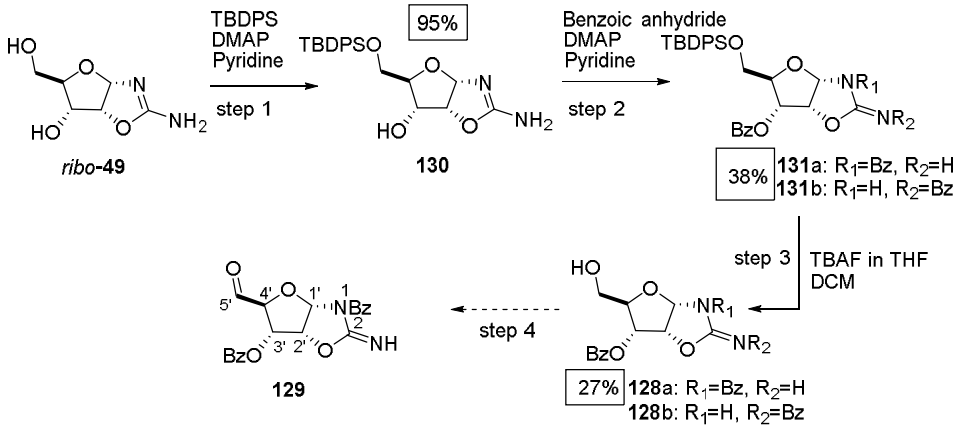


*Figure 7.2: ^{1}H NMR spectrum (600 MHz, D_2O, 6.00-3.50 ppm) of the oxidation of D-riboaminooxazoline (ribo-**49**), with TEMPO at 0°C for 7 h, according to the procedure described by Angelin et al.*[338]*, followed by lyophilisation and resuspension in D_2O. Spectrum shows incomplete reaction (20% of desired product 5'-oxo-D-ribofuranosyl aminooxazoline (**123**)).*

[xxx] Ongoing work in the Powner lab.

Taking into account the lack of purity afforded by this synthetic method and the difficulties in **123** purification in any amount necessary for future reactions and investigations (*vide supra*), we sought an alternative pathway to **123** that might ease large scale production and purification.

The polarity of 5′-oxo-aminooxazoline (**123**) being the cause of the aforementioned purification difficulties, an alternative pathway to **123** was sought through the use of protecting groups in order to ease purification. Consequently, we began to develop a four step strategy to C3′-benzoyl oxazoline **128** (*Scheme 7.14*). This scheme involved selective protection of C5′-hydroxyl moiety with tert-butyldiphenylsilyl ether (TBDPS), followed by perbenzoylation. Finally, deprotection of the primary C5′-hydroxyl and Swern oxidation would yield our desired protected 5′-oxo 3′-*O*-benzoyl-D-ribofuranosyl 1/2-*N*-benzoyl-aminooxazoline (**129**). In the interest of higher yield of desired key intermediate **119** (*vide supra*), we plan on carrying the protected form **129** through the reductive amination step and then performing the benzoyl group (Bz) deprotection to yield **119**. Furthermore, in order to compare the final deprotected **119** with one produced through aldehyde **123**, we will deprotect a portion of **129** to carry through the original route planned.



*Scheme 7.14: Proposed protecting strategy to yield a benzoylated form of 5'-oxo-riboaminooxazoline, **129**, in order to ease large scale production and purification. Isolated yield for each step performed is annotated with the corresponding molecule.*

A solution of D-riboaminooxazoline (*ribo*-**49**) was stirred with a slight excess of TBDPS-Cl and catalytic 4-dimethylaminopyridine (DMAP) in pyridine. After 20 h under a nitrogen atmosphere the reaction was quenched with methanol and concentrated *in vacuo*. ^{1}H NMR analysis showed near quantitative conversion of *ribo*-**49** to 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl aminooxazoline (**130**) (*Figure 7.3*). A small portion of the compound was purified further for characterisation, while the bulk of the reaction was carried into step 2 (*Scheme 7.14*, *vide supra*).

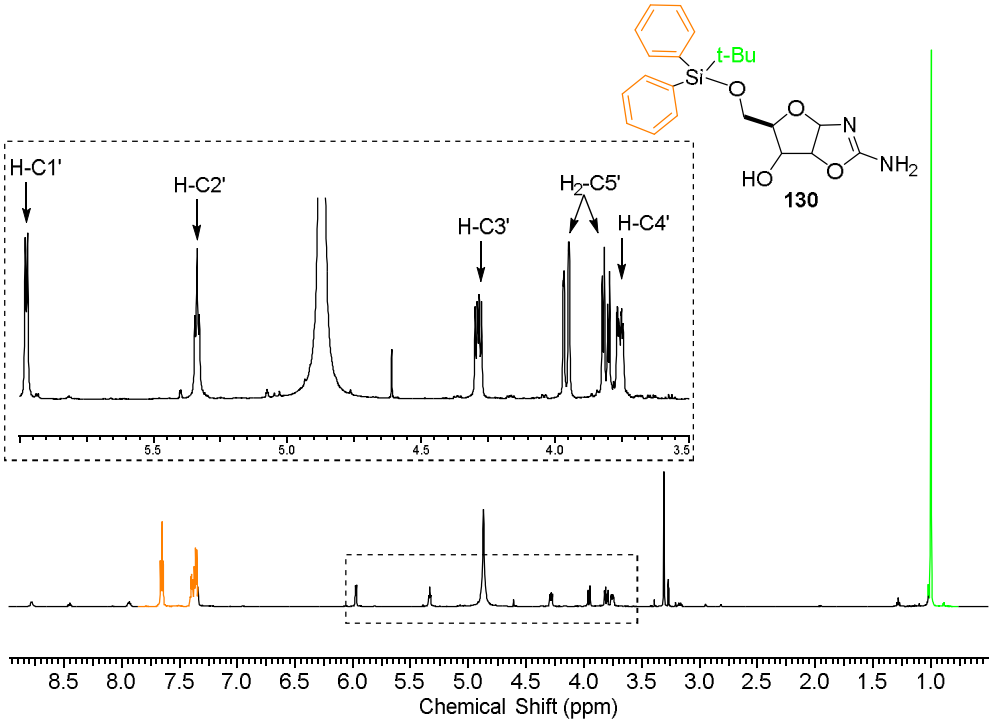


*Figure 7.3: ^{1}H NMR spectrum (600 MHz, CD_3OD, 9.00-0.50 ppm) of the reaction of D-riboaminooxazoline (ribo-**49**; 860mM) in Py with TBDPS-Cl (1.1 eq) and DMAP (cat. amount) after 20 h at RT. Solvent was removed in vacuo and residue analysed in CD_3OD. Inset (6.00-3.50 ppm): Details of the protons for the sugar moiety of product 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl aminooxazoline (**130**).*

Crude **130** was stirred in pyridine with benzoic anhydride (3 eq) and catalytic DMAP, and after 21 h TLC analysis indicated no residual **130**, so the solvent was removed *in vacuo*. The residue was then extracted into CH_2Cl_2 and purified by

FCC to yield 3′-*O*-benzoyl 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl 1/2-*N*-benzoyl-aminooxazoline (**131**; 38%). TLC analysis, after FCC, showed two spots, leading us to consider the presence of an alternative isomer **131**b (*Figure 7.4*), due to the presence of distinctive unprotected *N*1-H peaks in ^{1}H NMR spectrum (6.88 ppm, d, J = 9.8 Hz). A second FCC purification was fruitful in partially separating these two product: pure **131**a, confirmed by characterisation (*Figure 7.5, a.*), and a mixture of **131**a and **131**b (as interpreted by ^{1}H NMR analysis, *Figure 7.5, b.*) were isolated. **131**a and **131**b are thought to have formed by *N*1- and *N*2-benzoylation, respectively. However, if our interpretation is correct, once **131**a and **131**b are desilylated, they will both allow selective C5′-oxidation, and the mixture of isomers doesn't detract from moving forwards with the planned route.

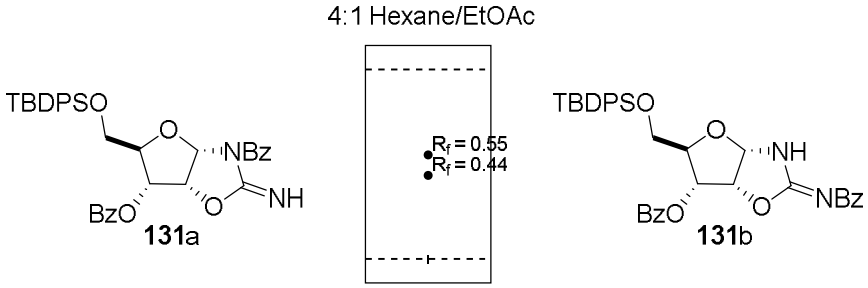


*Figure 7.4: TLC analysis of the reaction of 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl aminooxazoline (**130**; 109mM) in Py with benzoic anhydride (3 eq) and DMAP (cat.) after 21 h at RT, followed by FCC purification, showing the presence of two similar products, **131**a and **131**b (proposed structures of compounds **131**a and **131**b shown).*

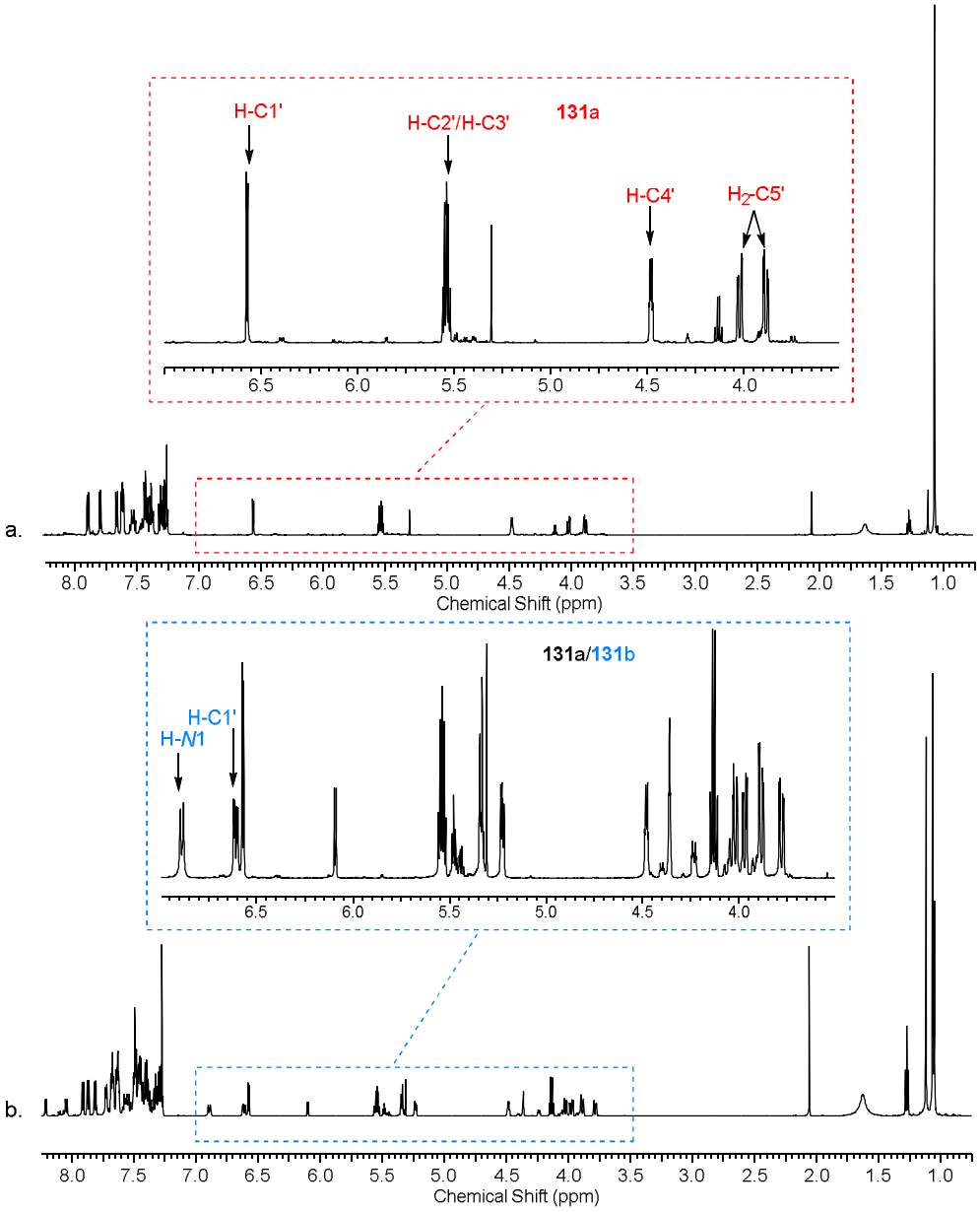


*Figure 7.5: ^{1}H NMR spectra (600 MHz, $CDCl_3$, 8.75-0.75 ppm) of the reaction of 5'-tert-butyldiphenylsilyl-ether-D-ribofuranosyl aminooxazoline (**130**; 109mM) in Py with benzoic anhydride (3 eq) and DMAP (cat.) after 21 h at RT, followed by two FCC purifications. a. Pure product 5'-tert-butyldiphenylsilyl-ether-D-ribofuranosyl 1-N-benzoyl-aminooxazoline (**131**a). Inset (7.00-3.50 ppm): Details of the protons for the sugar moiety of **131**a. b. Mixture of products **131**a and **131**b formed from N1 and N2-benzoylation respectively: **131**b assigned as far as possible. Inset (7.00-3.50 ppm): Details of the protons for the sugar moiety of **131**a and **131**b.*

The mixture of **131**a and **131**b was desilylated using tetra-n-butylammonium fluoride (TBAF). The reaction was stirred overnight, after which TLC indicated a complete reaction. FCC purification yielded 27% pure 3′-*O*-benzoyl-D-ribofuranosyl 1-*N*-benzoyl-aminooxazoline (**128**a; confirmed by NMR analysis, *Figure 7.6*, *a.*), as well as 35% of a mixture of **128**a and **128**b (again, likely the result of the same *N*1- and *N*2-benzoylation mentioned above, respectively, *Figure 7.6*, *b.*). We plan on oxidising **128** to form 5′-oxo-3′-*O*-benzoyl-D-ribofuranosyl 1-*N*-benzoyl-aminooxazoline (**129**) by Swern-type oxidation, in order to obtain our 5′-oxo protected species **129** (*Scheme 7.15*). This will allow to potentially access a benzoylated form of 5′-AICA riboaminooxazoline, **132**, through reductive amination (*vide infra*). Finally, benzoyl deprotection would afford key intermediate 5′-AICA riboaminooxazoline (**119**), which was the final intended product of this synthetic pathway.

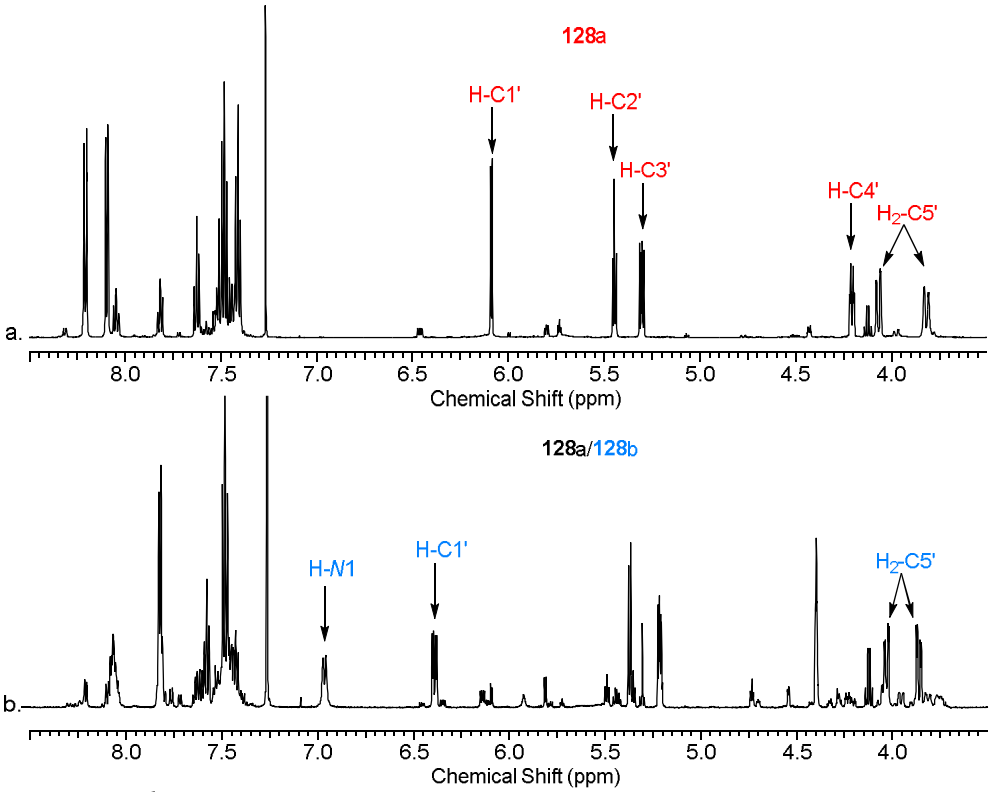
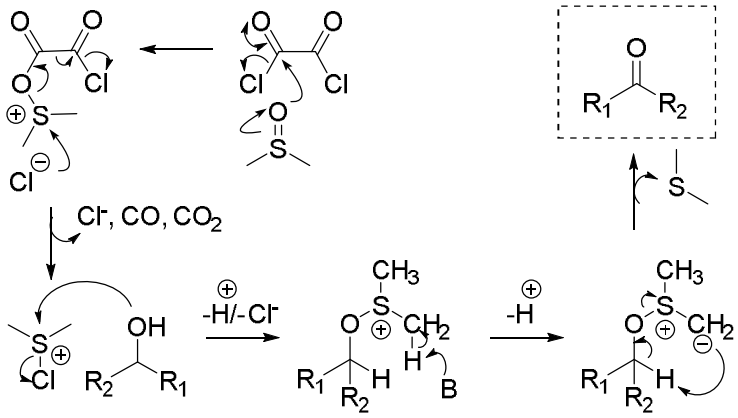


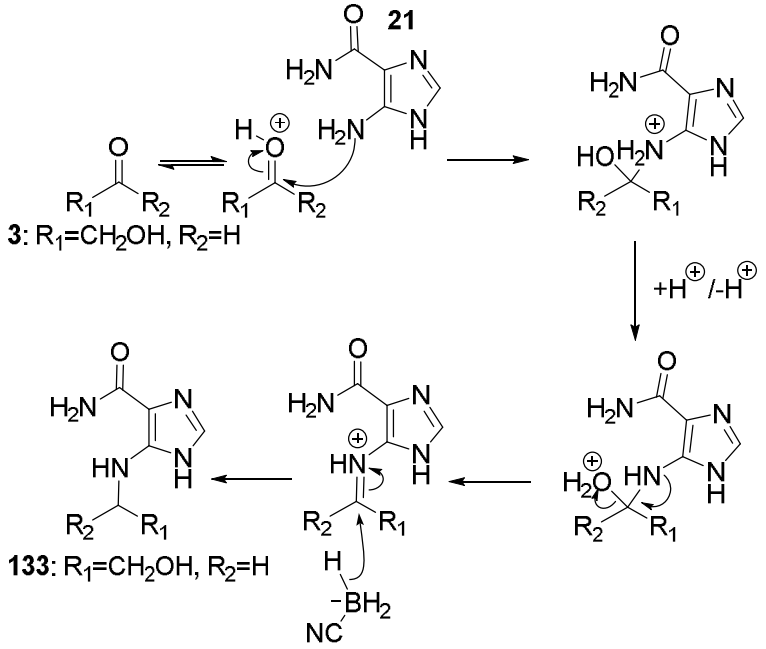
*Figure 7.6: ^{1}H NMR spectra (600 MHz, $CDCl_3$, 8.50-3.50 ppm) of the reaction of 5'-tert-butyldiphenylsilyl-ether-D-ribofuranosyl 1/2-N-benzoyl-aminooxazoline (**131**; 102mM) in DCM with TBAF (1 eq) after 18 h at RT, followed by FCC purification. a. Pure product 3'-O-benzoyl-D-ribofuranosyl 1-N-benzoyl-aminooxazoline (**128**a); b. Mixture of products **128**a and **128**b formed from N1 and N2-benzoylation respectively: **131**b assigned as far as possible.*



Scheme 7.15: Swern-oxidation mechanism using DMSO and oxalyl chloride, yielding the desired aldehyde/ketone (dotted box) from an alcohol (primary/secondary, respectively).

7.2.2. Synthesis of 5′-AICA riboaminooxazoline

Once produced in sufficient yield, 5′-oxoriboaminooxazoline (**123**), or its benzoylated form **129**, could then be reacted with AICA **21** (a purine precursor, *vide supra*) under reductive amination conditions in order to access intermediate 5′-AICA riboaminooxazoline (**119**), or the benzoylated form **132**, respectively (in which case a further deprotection step would be needed to yield the intermediate **119** we require for our intended prebiotic investigations). Before performing this reaction with aldehyde **123**, various proof of concept reductive amination reactions (*Scheme 7.16*) were performed, with different starting materials and conditions, in order to find suitable conditions to investigate the reductive amination of valuable aldehyde **123**.



*Scheme 7.16: Mechanism of reductive amination using sodium cyanoborohydride (**124**) as exemplified by AICA **21**.*

First the reaction of glycolaldehyde (**3**; 100mM)[xxxi] with AICA **21** (1-3 eq) in the presence of sodium cyanoborohydride (**124**) was studied between pH 4 and 9 in order to find optimal conditions. It was apparent from the reactions that a lower pH (pH 4) led to a more efficient reaction (*Figure 7.7*, *a.*), with greater yield of product 2-hydroxyethyl-AICA (**133**) compared to side-products (ethylene glycol (**134**); formed as the reduction product of aldehyde **3**, and confirmed by sample spiking, *Figure 7.8*, *b.*; and 2,3-dihydroxypropanenitrile (**62**); suggested by-product formed from the action of trace amounts of cyanide on aldehyde **3**, based on ^{1}H NMR analysis). This higher selectivity at lower pH is most likely due to the greater proportion of imminum ion formed (*Scheme 7.16*, *vide supra*), leading to higher rate of reduction.

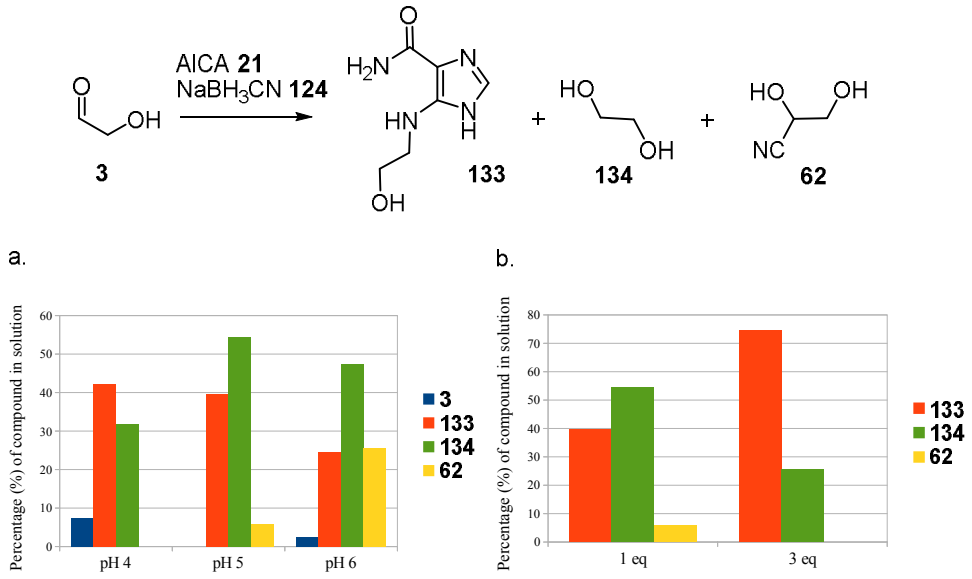


*Figure 7.7: Reductive amination of glycolaldehyde (**3**; 100mM) and AICA **21**, with sodium cyanoborohydride (**124**), in water and at RT. All analysis performed after 1 h. a. Comparison of pH on yield of reductive amination product 2-hydroxyethyl-AICA (**133**). All reactions were performed with 1:1:3 ratio of **3**, **21** and **124** respectively. b. Comparison of the effect of equivalents of **21** on yield. All reactions were done at pH 5.*

[xxxi] Glycolaldehyde (**3**) was used a generic reagent for the initial reductive amination studies because it is a readily available sugar synthon.

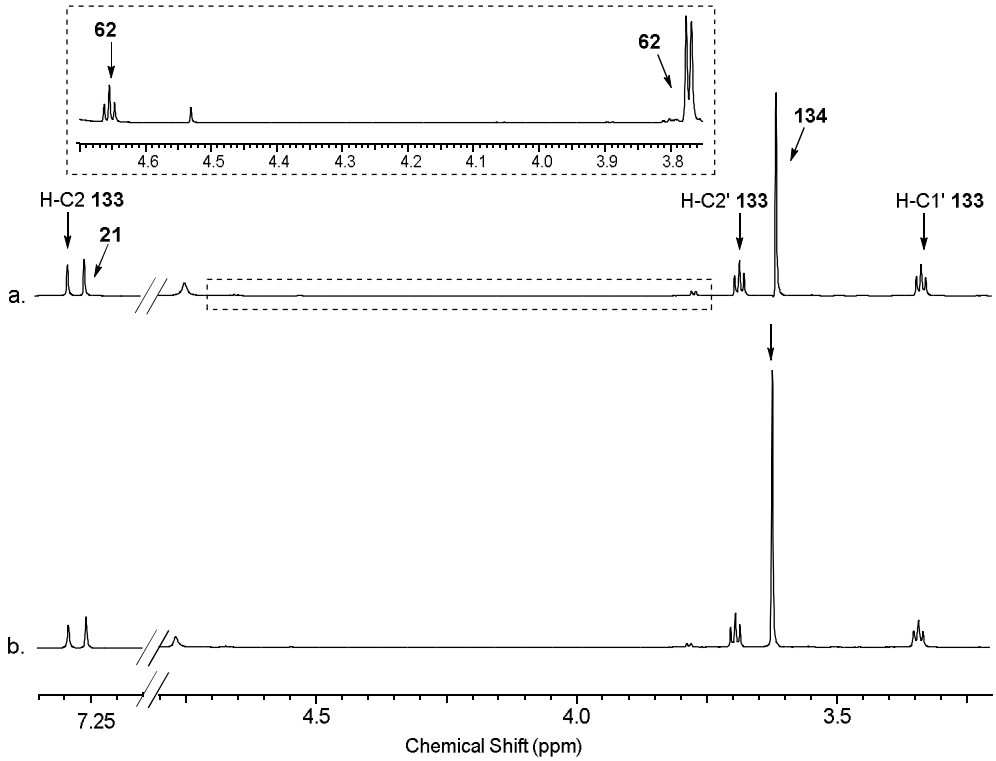
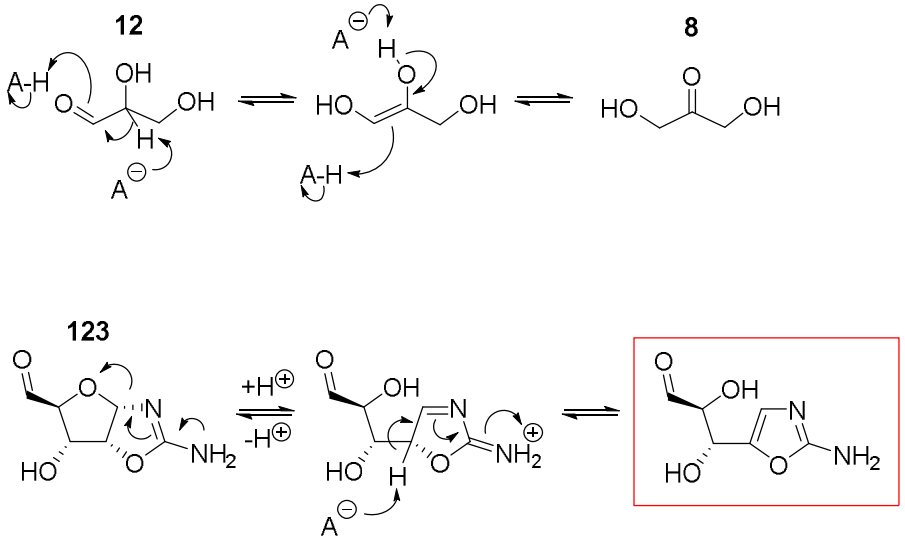


Figure 7.8: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 7.35-3.20 ppm) of the reductive amination of glycolaldehyde (**3**; *100mM) and AICA **21** (1 eq) with sodium cyanoborohydride (**124**; 3 eq), at pH 6, and at RT. Showing formation of desired product 2-hydroxyethyl-AICA (**133**) alongside by-products ethylene glycol (**134**) and 2,3-dihydroxypropanenitrile (**62**). a. Crude reaction after 2.5 h. b. spiked with **134**. Inset (4.70-3.75 ppm): details of **62** peaks.*

These experiments were performed with the reductive amination of 5′-oxo riboaminooxazoline (**123**) in mind and thus we wanted to find a pH that would give us the best yield without affecting the integrity of the aminooxazoline ring of **123** (pK_a 6.54[312]) or the product formed. The reaction done at pH 4 gave a higher yield of product **133** (42% *vs* 40%, pH 4 and 5 respectively) over side products **134** and **62** (32% *vs* 60% combined yields, pH 4 and 5 respectively; based on ^{1}H NMR analysis). However, investigation of the effect of AICA **21** concentration on yield at milder pH 5 was carried out in order to avoid pH conditions too far from the pK_a of the aminooxazoline ring of **123**. ^{1}H NMR analysis of this study showed a clear preference for product **133** formation (over reduction of the aldehyde **3** to **134**) when more **21** was added.

Indeed, with 3 eq of **21** the reaction was found to produce a crude yield of **133** of 75% *vs* the 40% observed when 1 eq is used (*Figure 7.7*, *b.*), and no starting material or side product **62** was observed.

The results obtained with glycolaldehyde (**3**) encouraged us to repeat these sets of experimental conditions on glyceraldehyde (**12**), which is capable of undergoing Lobry de Bruyn-van Ekenstein transformation to dihydroxyacetone (**8**), and therefore is more representative of our aldehyde **123** which can also undergo a similar isomerisation through its open-chain version (drawn in red box, *Scheme 7.17*).



*Scheme 7.17: Mechanism of Lobry de Bruyn-van Ekenstein transformation as exemplified by glyceraldehyde (**12**) to dihydroxyacetone (**8**) (top). Mechanism of aminooxazoline **123** opening to yield a species capable of aldose/ketose isomerisation (red box).*

Therefore, we wanted to verify that reductive amination to yield the correct product (through aldehyde amination) was selective over any other pathways (aldose/ketose isomerisation and reduction of the starting material), when starting with glyceraldehyde (**12**). As with the results obtained with C_2 aldehyde **3** (*vide supra*), lower pH (pH 3) showed more selective formation of desired product 2,3-dihydroxypropyl-AICA (**135**) compared to higher pH, with higher yields of **135** (69% *vs* 29%, pH 3 and pH 7 respectively). Additionally, less by-products,

glycerol (**67**) and 2,3,4-trihydroxybutanenitrile (**63**) (analogous to **134** and **62** described above) were observed from competing chemistry (31% *vs* 71% combined yields, pH 3 and pH 7 respectively; *Figure 7.9*, *a*.). Again we compared the effect of the ratio of starting materials **12** and AICA **21**, and found that a large excess of **21** (6 eq) led to highest yields of desired product **135** at pH 5[xxxii] (68% **135**), while rendering the reaction more selective towards the desired pathway (reductive amination of aldehyde **12**) as no by-product **63** was observed (*Figure 7.9, b.*). No other species were detected (*Figure 7.10*), suggesting no isomerisation of **12** to dihydroxyacetone (**8**) occurred in any significant amount, and no further reaction, such as reductive amination, occurred on **8**.

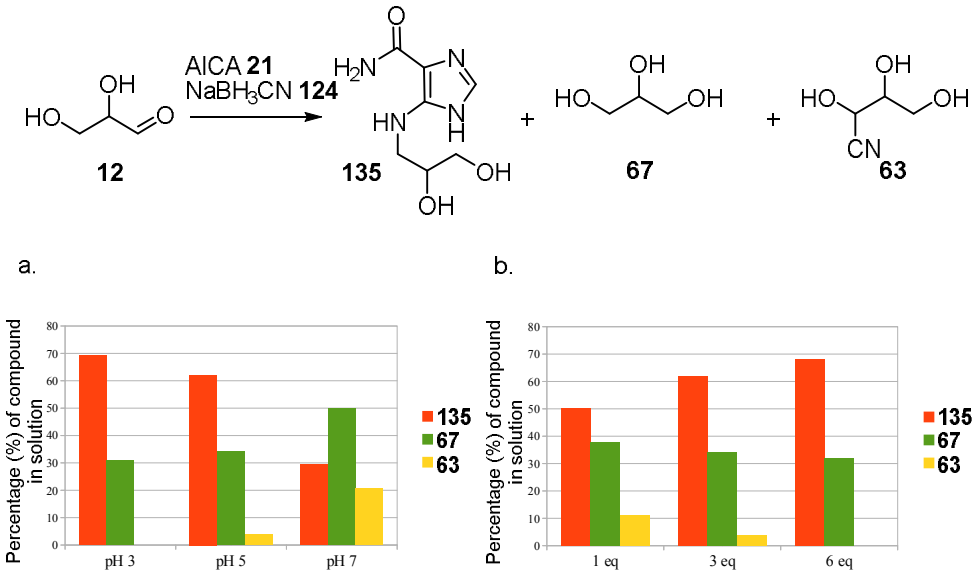


*Figure 7.9: Reductive amination of glyceraldehyde (**12**; 100mM) and AICA **21**, with sodium cyanoborohydride (**124**), in water and at RT. All analysis performed after 1 h to 2.5 h. a. Comparison of pH on yield of reductive amination product 2,3-dihydroxypropyl-AICA (**135**). All reactions were performed with 1:3:3 ratio of **12**, **21** and **124** respectively. b. Comparison of the effect of equivalents of **21** on yield. All reactions were done at pH 5.*

[xxxii] We chose to study the effect of ratios at pH 5, for the same reason as given above for **3**: avoiding pH conditions too far from the pK_a of the aminooxazoline ring of **123** (pK_a 6.54).

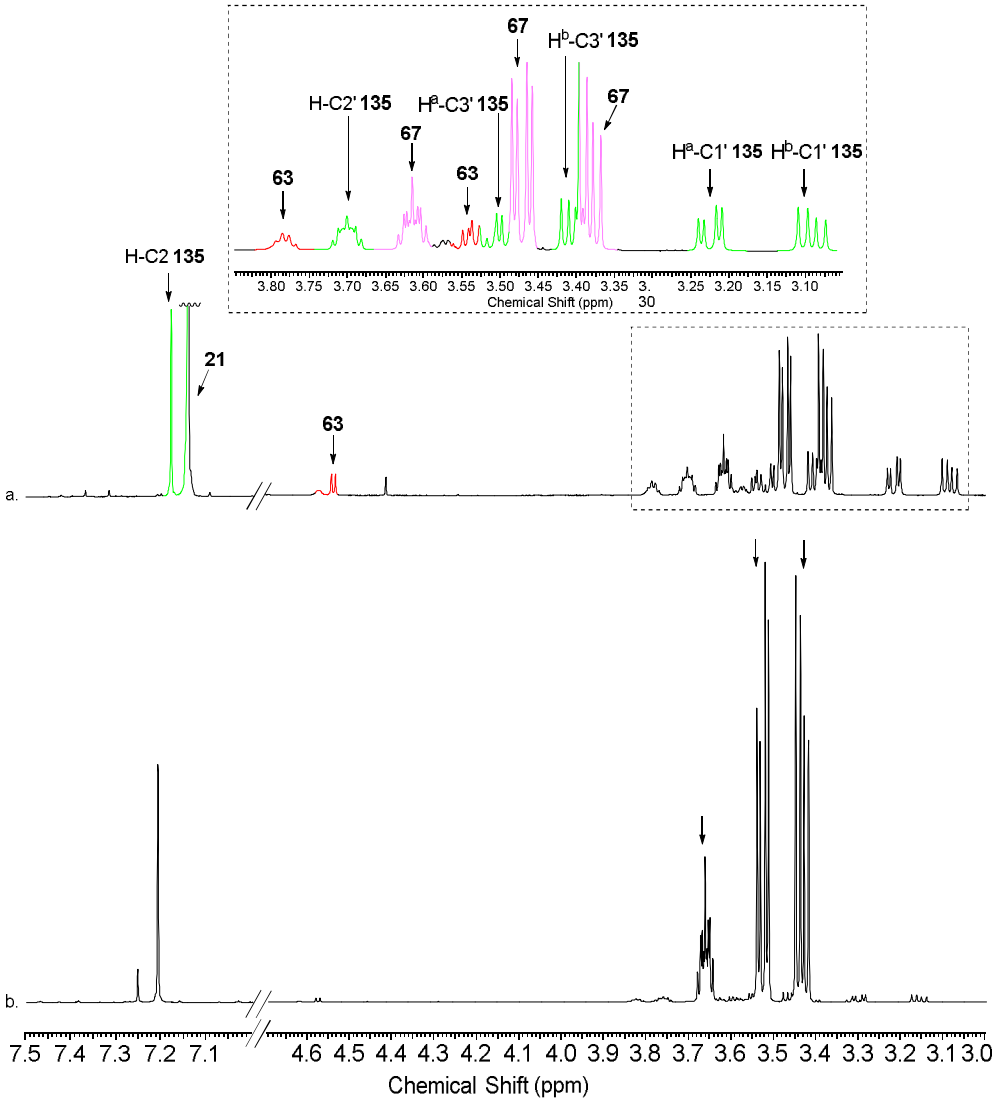


*Figure 7.10: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 7.50-3.00 ppm) of the reductive amination of glyceraldehyde (**12**; 100mM) and AICA **21** (3 eq) with sodium cyanoborohydride (**124**; 3 eq), at pH 7 in 0.8M P_i buffer, and at RT. Showing formation of desired product 2,3-dihydroxypropyl-AICA (**135**) alongside by-products glycerol (**67**) and 2,3,4-trihydroxybutanenitrile (**63**). a. Crude reaction after 5 h. b. spiked with **67**. Inset (3.85-3.05 ppm): details of peaks and assignments for **135**.*

2D NMR analysis confirmed the structure of our reductive amination product **135** through clear ^{1}H-^{13}C HMBC correlation between C5 of the AICA moiety and H-C3′ (*Figure 7.11*). We were confident that the reductive amination took place on the correct nitrogen, *N*-C5, due to lack of correlation between H-C3′ and C2

(and C3′ and H-C2). The preference for this nitrogen over other AICA nitrogens can be rationalised by the fact that it is the most nucleophilic one. Indeed, the lone pairs of *N*1 and $\underline{N}H_2CO$-C4 are delocalised (in the aromatic imidazole ring and the carbonyl group respectively). Furthermore, any reaction with *N*3 would be reversible as it would form an unstable iminium that could only be stabilised by the unfavourable breaking of the aromaticity of the imidazole ring.

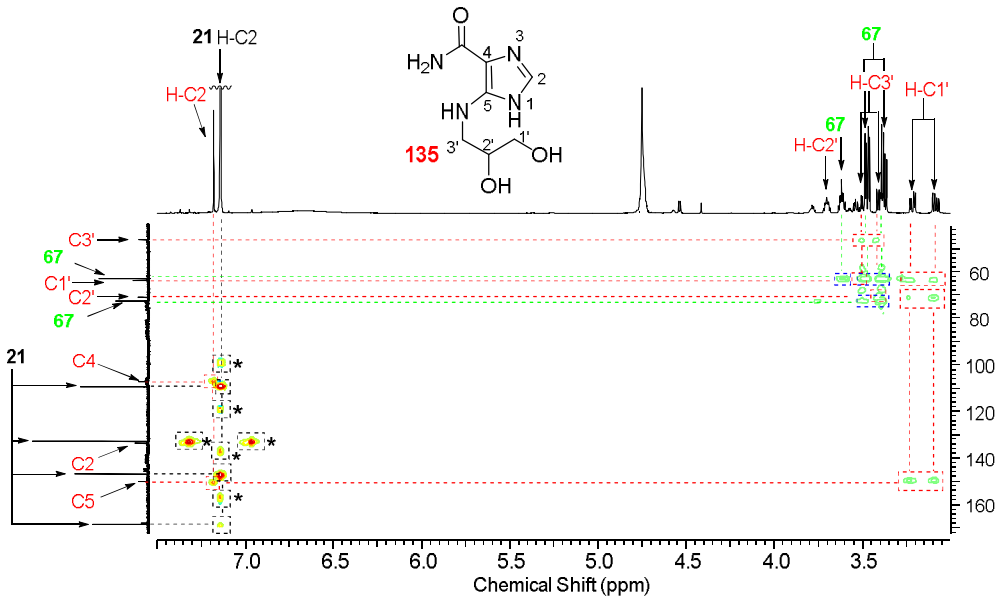


*Figure 7.11 :^{1}H-^{13}C HMBC (600 MHz, H_2O/D_2O 9:1) of the reductive amination of glyceraldehyde (**12**; 100mM) and AICA **21** (3 eq) with sodium cyanoborohydride (**124**; 3 eq), at pH 7 in 0.8M P_i buffer, after 5 h at RT. Showing formation of desired product 2,3-dihydroxypropyl-AICA (**135**; evidenced by the coupling between the C5 of the AICA moiety, and the H-C1' of the sugar moiety), alongside by-product glycerol (**67**). Cross coupling of **21** annotated by *.*

Exploration of the reductive amination of glyceraldehyde (**12**) with AICA **21** confirmed that this method could be used on a more complex aldehyde capable of aldose/ketose isomerisation. Furthermore, we demonstrated efficient, rapid and selective reductive amination of **12** (our model aldehyde) and **21** (our desired amine), with only minor amounts of reduction of the aldehyde when using optimised conditions of 1:6:3 aldehyde:**21**:**124** at pH 5 at RT.

Based on these results found by investigation of reductive amination of aldehydes **3** and **12**, we finally investigated the reductive amination of 5′-oxo riboaminooxazoline (**123**), with **21** (6 eq), and **124** (3 eq) at pH 5. Acetate buffer (1M) at pH 4.7 was used in order to stabilise the pH, and the reaction was monitored over 3 h. After 1 h, full NMR analysis showed complete consumption of starting material **123**, and showed the presence of the desired product 5′-AICA riboaminooxazoline (**119**; 36% based on ^{1}H NMR integration after 1 h, confirmed through HMBC correlation between H-C5′ and C5″ (*Figure 7.12*). We also observed formation of riboaminooxazoline (*ribo*-**49**; 51%), the by-product formed from reduction of the aldehyde **123** (confirmed by sample spiking, *Figure 7.13*, *b.*).

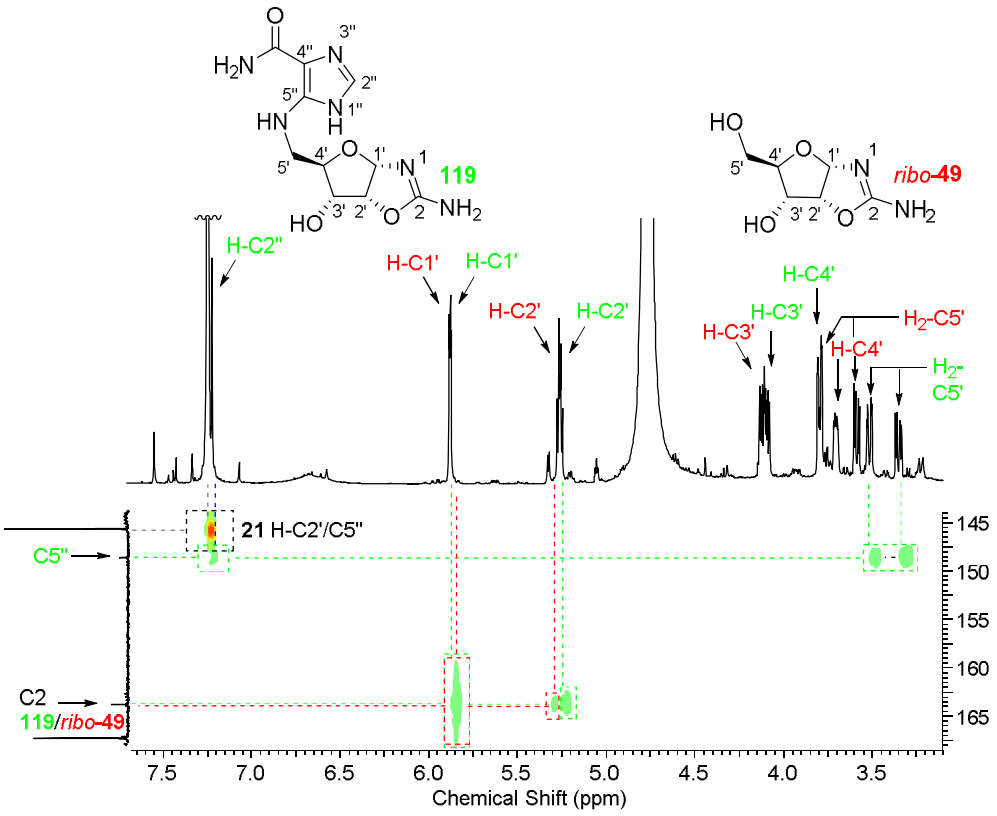


*Figure 7.12: ^{1}H-^{13}C HMBC (600 MHz, H_2O/D_2O 9:1) of the reductive amination of 5'-oxo-riboaminooxazoline (**123**; 100mM) and AICA **21** (6 eq) with sodium cyanoborohydride (**124**; 3 eq), in acetate buffer (1M, pH 4.7), after 3 h at RT. Showing formation of desired product 5'-AICA-riboaminooxazoline (**119**) (evidenced by the coupling between the C5" of the AICA moiety, and the H-C5' of the sugar moiety), alongside by-product riboaminooxazoline (ribo-**49**).*

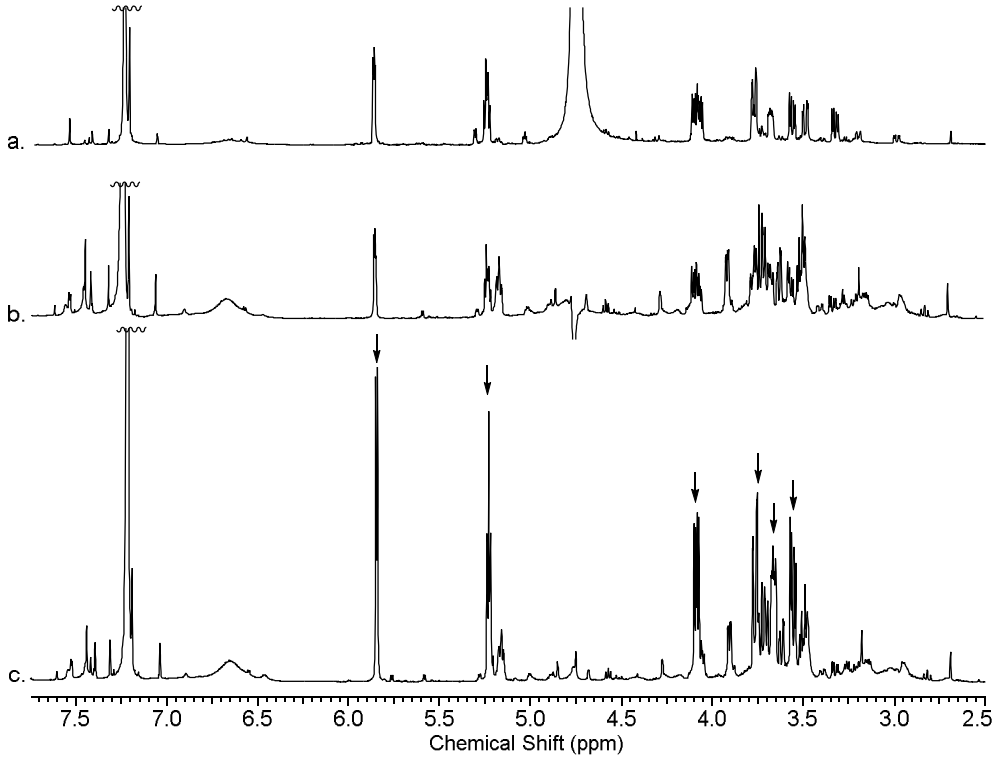


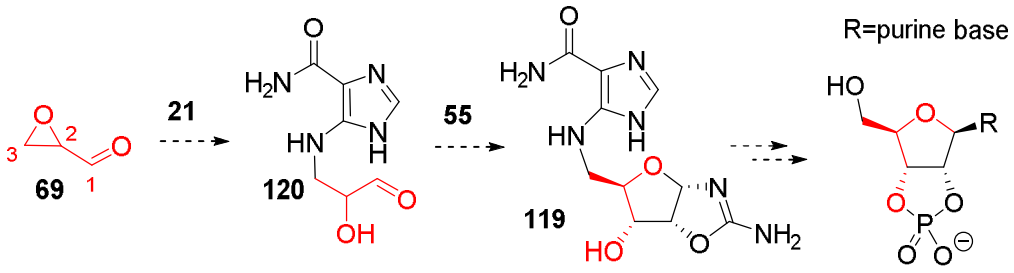
*Figure 7.13: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 7.75-2.50 ppm) of the reductive amination of 5′-oxo-riboaminooxazoline (**123**; 100mM) and AICA **21** (6 eq) with sodium cyanoborohydride (**124**; 3 eq), in acetate buffer (1M, pH 4.7), at RT. Showing formation of desired product 5′-AICA-riboaminooxazoline (**119**) alongside by-product riboaminooxazoline (ribo-**49**). a. Crude reaction after 3 h. b. crude reaction after a further 3 weeks. Shift in peaks confirms presence of two species that were partially overlapped. c. spiked with ribo-**49**.*

This small scale study showed promising results concerning the synthetic formation of key intermediate 5′-AICA-riboaminooxazoline (**119**). Purification and optimisation of this reaction will be carried out once an optimised route to pure 5′-oxoriboaminooxazoline (**123**) has been found. After optimisation of the synthesis leading to **119**, we will study prebiotic pathways to regioselective glycosidation as discussed above. Importantly, a prebiotically plausible pathway to key intermediate **119** also needs to be established in order for **119**, and its subsequent reactions, to have prebiotic significance, which is something we will explore in more detail in *Section 7.3.2* (*vide infra*).

7.3. Prebiotic 5′-substituted aminooxazolines

7.3.1. Nucleobase precursors

In previous sections, we made the case for the use of epoxide glycidaldehyde (**69**) in prebiotic chemistry, by demonstrating it can be used to regiospecifically incorporate phosphate by nucleophilic attack on the epoxide moiety both in its free form (forming glyceraldehyde-3-phosphate (**68**), *Section 3*), and when the aldehyde moiety reacted with 2-aminooxazole (**55**) to form intermediate oxazolines **97** (leading to 5′-phosphate aminooxazoline (**53**) formation). We plan on exploiting **69**'s potential for efficient nucleophilic attack on the C3 position in order to potentially tether AICA **21** (one of the most robust and widely reported purine precursors under investigation in prebiotic chemistry, *vide supra*) to this glyceraldehyde analogue, forming glyceraldehyde-3-AICA (**120**; *Scheme 7.18*). We could then use this new species in the same way glyceraldehyde (**12**) is used in the synthesis of aminooxazolines (*vide supra*) and thereby obtain the sought key intermediate 5′-AICA riboaminooxazoline (**119**) prebiotically (*Scheme 7.18*).



*Scheme 7.18: Proposed route to purine ribonucleotides through the reaction of glycidaldehyde (**69**) and purine precursor AICA **21**, followed by reaction with 2-aminooxazole (**55**) to obtain key intermediate 5′-AICA riboaminooxazoline (**119**).*

We investigated the reaction of **69** (140mM) with **21** (3 eq) in water, at RT and at a range of pH. At pH 4.4 and 7.7, **69** was completely consumed after 24 h with the formation of a new product in 56% and 74% crude yield, respectively (based upon analysis of ^{1}H NMR spectra compared to other species present in solution).

Careful analysis of the compound by 2D-NMR allowed cyclic hemi-aminal imidazole **136** to be determined as the major product due to the HMBC correlation of peaks H-C1′ and H-C3′ with C5, H-C3′ with C2, and the reciprocal correlation between C3′ (and C2′) with H-C2 (two anomers are present in 1:4 ratio, only major anomer is labelled, *Figure 7.14*). The lack of correlation between C2 and H-C1′ (and the reciprocal H-C2 with C1′) gave us confidence that we were observing formation of product **136** through the initial attack of $\underline{N}H_2$-C5 onto the aldehyde moiety of **69** followed by cyclisation onto N1 (*Scheme 7.19*).

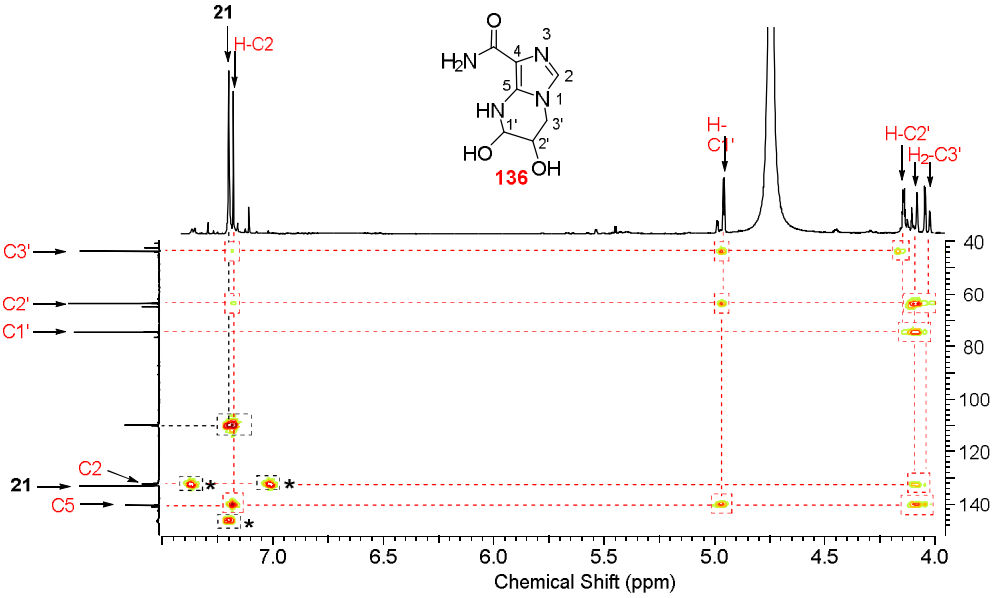
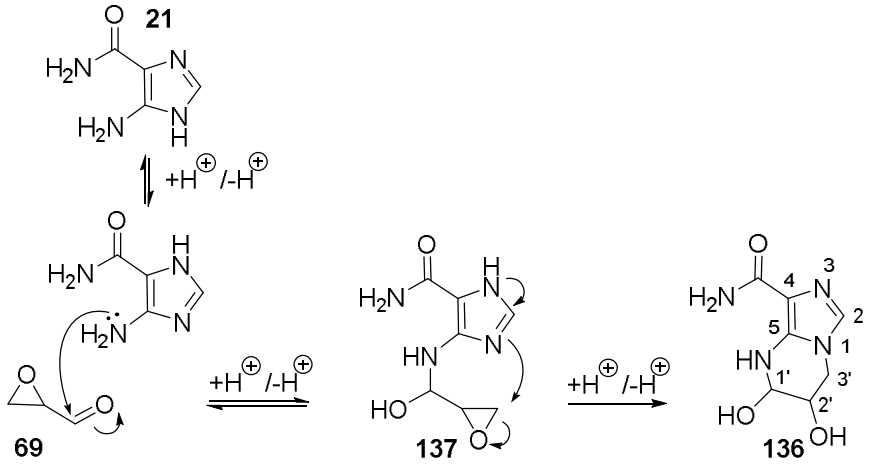
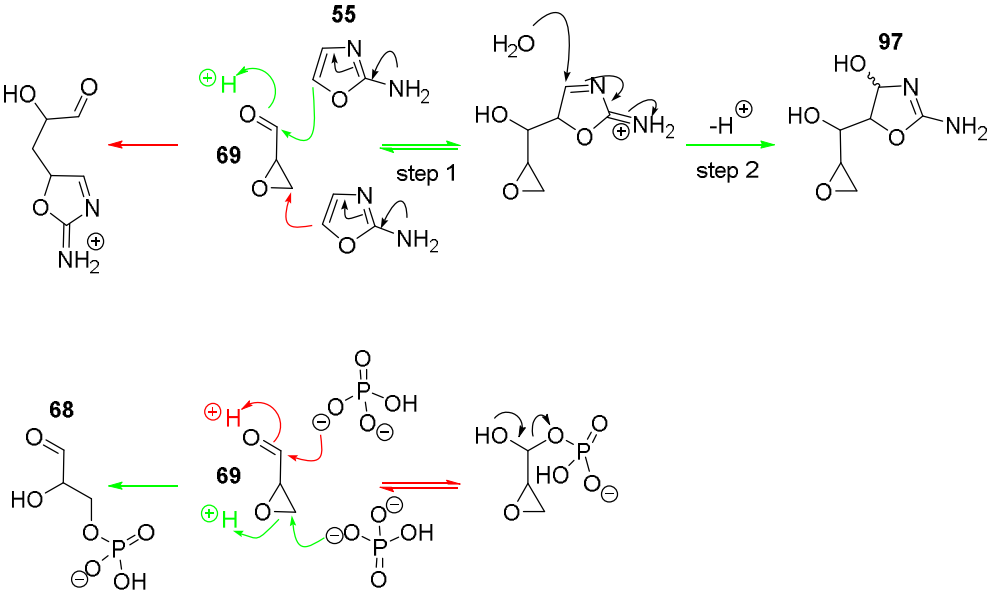


*Figure 7.14: ^{1}H-^{13}C HMBC (600 MHz, H_2O/D_2O 9:1) of the reaction of glycidaldehyde (**69**; 140mM) with AICA **21** (3 eq) in H_2O/D_2O (9:1) at pH 7.7 and at RT. Showing formation of cyclic hemi-aminal imidazole **136** (evidenced by the coupling between the C5 of the AICA moiety, and the H-C1' and H-C3' of the sugar moiety and between C2 and H-C1' and H-C3'). Cross coupling of **21** annotated by *. Only major anomer of **136** is labelled.*



*Scheme 7.19: Proposed mechanism of the reaction of glycidaldehyde (**69**) with AICA **21** leading to the formation of cyclic hemi-aminal imidazole **136**.*

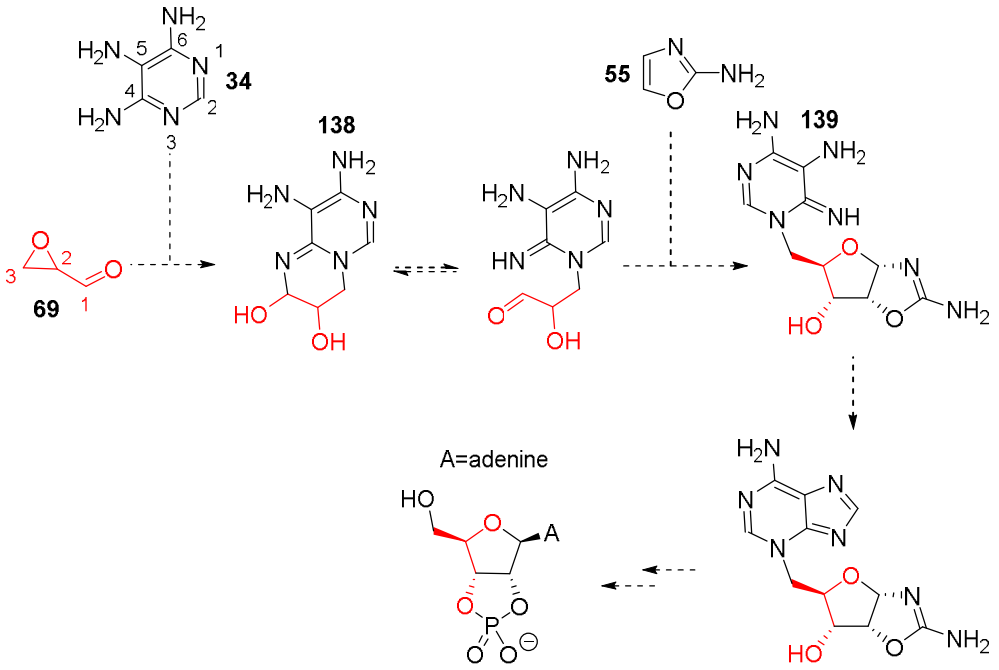
Though imidazole **136** is not our desired product glyceraldehyde-3-AICA (**120**), its synthesis is in no way a surprise; we anticipated that rapid equilibration of amino-imidazole **21** and **69** would provide hemi-aminal **137**, followed by intramolecular rearrangement to yield **136** (*Scheme 7.19*). In our reductive amination experiments described above, we also see preferential and rapid attack of the $\underline{N}H_2$-C5 nitrogen onto the aldehyde (of glyceraldehyde (**12**) and riboaminooxazoline (*ribo*-**49**)), due to the increased nucleophilicity of that nitrogen compared to other ones present in AICA **21** (rational for this greater nucleophilicity is detailed in *Section 7.2.2.*). Our experiments of **69** and 2-aminooxazole (**55**) described in *Section 4* showed synthesis of oxazolines **97** from the attack of **55** onto the aldehyde moiety of **69** (*Scheme 7.20*, step 1, reversible), followed by trapping of the adduct formed by addition of water (*Scheme 7.20*, step 2, irreversible). This mechanism is similar to the one observed in the reaction of epoxide **69** and AICA **21**, although the proximity of the *N*1 nucleophile results in cyclisation in this case (*Scheme 7.19*, *vide supra*). These are different to the preference observed in the reaction of **69** and P_i described in *Section 3* where phosphate nucleophilically attacks the epoxide moiety of **69** forming glyceraldehyde-3-phosphate (**68**; *Scheme 7.20*). Unlike the reaction with **55**, the addition of phosphate to the aldehyde moiety of **69** is reversible, as P_i is a good leaving group, and as such this pathway is not observed in the reaction of **69** and P_i (*Scheme 7.20*).



*Scheme 7.20: Comparison of the preferred reaction of nucleophiles on glycidaldehyde (**69**): aldehyde vs epoxide attack. Top: Reaction of **69** with 2-aminooxazole (**55**) yielding oxazoline **97**. Observed attack on the aldehyde moiety of **69** (green) over attack on the epoxide moiety (not observed; red). Bottom: Reaction of **69** with P_i yielding glyceraldehyde-3-phosphate (**68**). Observed attack on the epoxide moiety of **69** (green) over reversible attack on the aldehyde moiety (not observed; red).*

The observation that the nitrogenous nucleophiles used preferentially react with the aldehyde moiety of **69**, as opposed to the epoxide moiety as we had hoped, led us to envisage a new pathway to purine ribonucleotides, *via* 4,5,6-triaminopyrimidine (**34**), that made use of this preference (*Scheme 7.21*). If the exocyclic cyclic nitrogens $\underline{N}H_2$-C4[xxxiii] of **34** attack the aldehyde of **69** in an analogous manner to $\underline{N}H_2$-C5 of **21**, then we could obtain cyclic hemi-aminal **138**, which could lead us to a new key intermediate 5′-triaminopyrimidine riboaminooxazoline (**139**), through ring opening and reaction with 2-aminooxazole (**55**). This key intermediate **139**, would present the correct purine-nitrogen to the anomeric position in a similar way to 5′-AICA riboaminooxazoline (**119**). Furthermore, the C2-symmetery of **34** presented a highly interesting opportunity to maximise the selectivity of said tethering.

[xxxiii] Exocyclic nitrogens at the C4 and C6 position are identical due to symmetry at C2 in **34**. Therefore, $\underline{N}H_2$-C4 and $\underline{N}H_2$-C6 are interchangeable.



*Scheme 7.21: Proposed route to purine ribonucleotides through the reaction of glycidaldehyde (**69**) and adenine precursor 4,5,6-triaminopyrimidine (**34**), followed by reaction with 2-aminooxazole (**55**) to obtain new key intermediate 5′-triaminopyrimidine riboaminooxazoline (**139**).*

A time-course reaction of **69** and **34** was carried out at RT. Initial results indicated a quick reaction (only 1.5% **69** left after 15 min, relative to other species in solution) to form a new product triaminopyrimidine **140**. Interestingly the product form showed correlation between H-C1′/H-C2′ and C6 by ^{1}H-^{13}C HMBC NMR analysis (*Figure 7.15*), but no correlation between H-C3′ and C6. Additionally, the pyrimidine ring proton H-C2 correlated with C2′. These 2D-NMR data suggest that a branched cyclic product, such as **140**, was formed (as depicted in *Figure 7.15*). Furthermore, we observe a NOESY correlation between H-C2 and C2′/C3′ but not C1′, suggesting that the branched portion of the ring was proximal to the C2 position (as depicted in *Figure 7.15*).

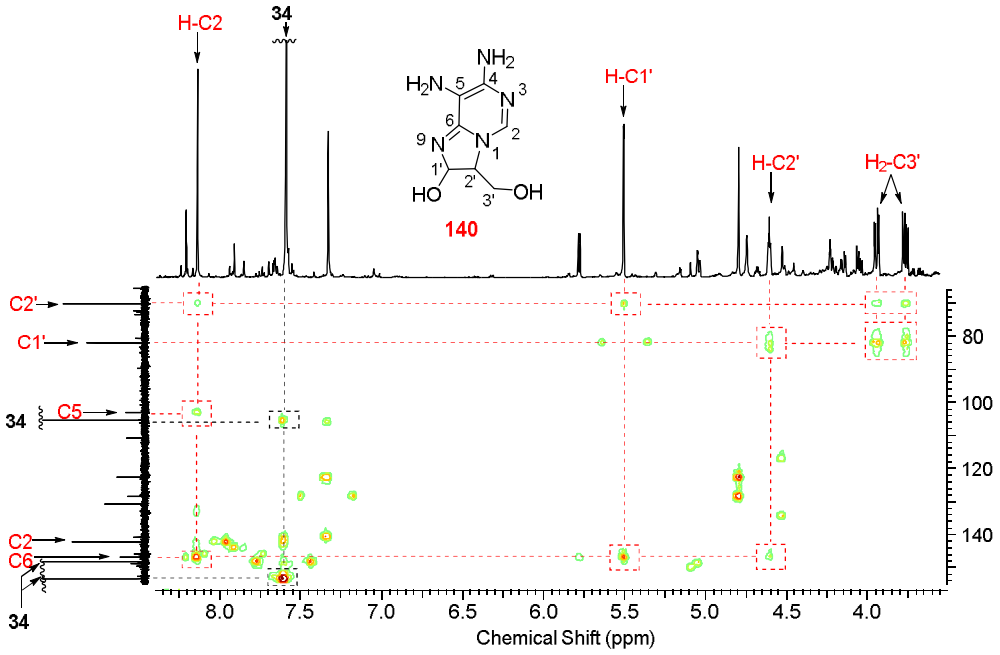
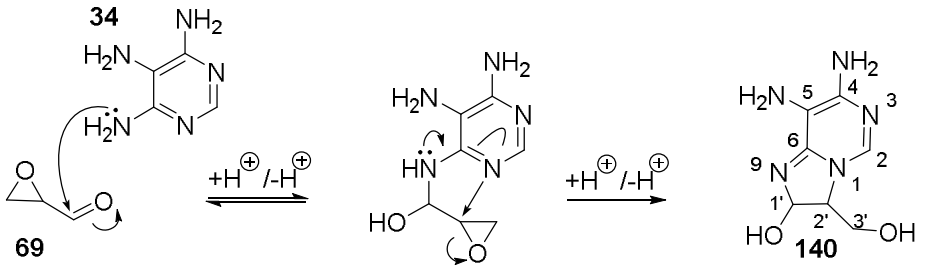


*Figure 7.15: ^{1}H-^{13}C HMBC (600 MHz, H_2O/D_2O 9:1) of the reaction of glycidaldehyde (**69**; 140mM) with 4,5,6-triaminopyrimidine (**34**; 3 eq) in H_2O/D_2O (9:1) at pH 7 and at RT. Showing formation of branched cyclic hemi-aminal **140** (evidenced by the coupling between the C6 of the pyrimidine moiety, and the H-C1' and H-C2' of the sugar moiety and between H-C2 and C2'). Only the major anomer of **140** is labelled.*

We postulated that once again the initial attack occurred from a primary amine onto the aldehyde moiety (*Scheme 7.22*), consistent with the reactions of glycidaldehyde (**69**) with AICA **21** (*Scheme 7.19*) and 2-aminooxazole (**55**) (*Scheme 7.20*, *vide supra*). Due to the constitutional similarities between **69** and glyceraldehyde (**12**), we can also draw parallels with **12** reactions with nitrogenous nucleophiles: in the reductive amination of **12** and **21** (*Section 7.2.2*) we observed selective reductive amination on the *N*-C5, which we explained by the theoretical higher nucleophilicity compared to other present nitrogens. Similarly, in the screens of multicomponent reaction of various aldehydes, including glycolaldehyde (**3**) and **12**, with **21** and **55**, Powner *et al.*[274] generate potential purine precursors *via* the initial selective attack of the same exocyclic *N*-C5 of **21** onto the aldehydes (*Scheme 7.5*, *vide supra*). This suggested that, as hypothesised in the reductive amination studies above, the exocyclic primary

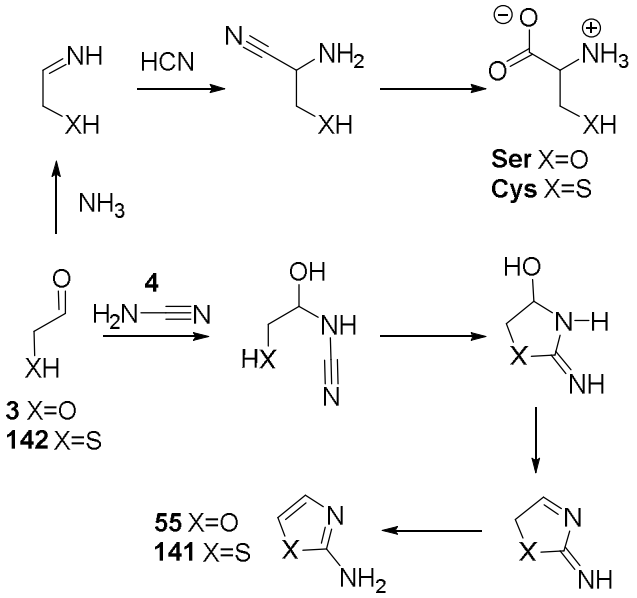
amine *N*-C5 of AICA **21** was more nucleophilic. Therefore, that the analogous exocyclic nitrogen *N*-C6 of triaminopyrimidine **34** should react with the aldehyde of **69** was expected and unsurprising (*Scheme 7.22*). What was unexcepted was that the subsequent cyclisation would happen on the C2 of **69**, yielding branched product **140**. Even though it has been suggested that α,β-epoxide carbonyls should preferentially be nucleophilically attacked at the α-position[303] (due to increased electrophilicity at this carbon), we have so far only observed nucleophilic substitution at the β (C3) position of **69** or an adduct of **69** (such as **97**; *Scheme 7.20*). In the case of hemi-aminal formation, one might expect a similar increased electrophilicity at the α-position, which was observed solely in the reaction of **69** and triaminopyrimidine **34**. There is some literature precedent for the formation of a similar branched cyclic hemi-aminal in the reaction **69** and deoxyguanosine (studied in the context of the carcinogenic nature of **69**), although the authors report a mixture of the 5-membered ring (analogous to **140**) and the 6-membered ring (analogous to **138**)[339].



*Scheme 7.22: Proposed mechanism of the reaction of glycidaldehyde (**69**) with 4,5,6 triaminopyrimidine (**34**) leading to the formation of branched cyclic hemi-aminal **140**.*

In this section so far we have investigated ways of tethering purine precursors to epoxide **69**, without the desired outcomes. Our ultimate goal in the research we have described has been to find pathways or key molecules to lead to purine and pyrimidine ribonucleotide synthesis. We have approached the subject of other class of metabolites in *Section 2*, namely proteins (through amino acid synthesis), however we have not yet looked at the synthesis of another important biomolecule: DNA. The central role of DNA in life is clearly undisputed, and as such has been the subject of prebiotic chemistry research. One intriguing possible

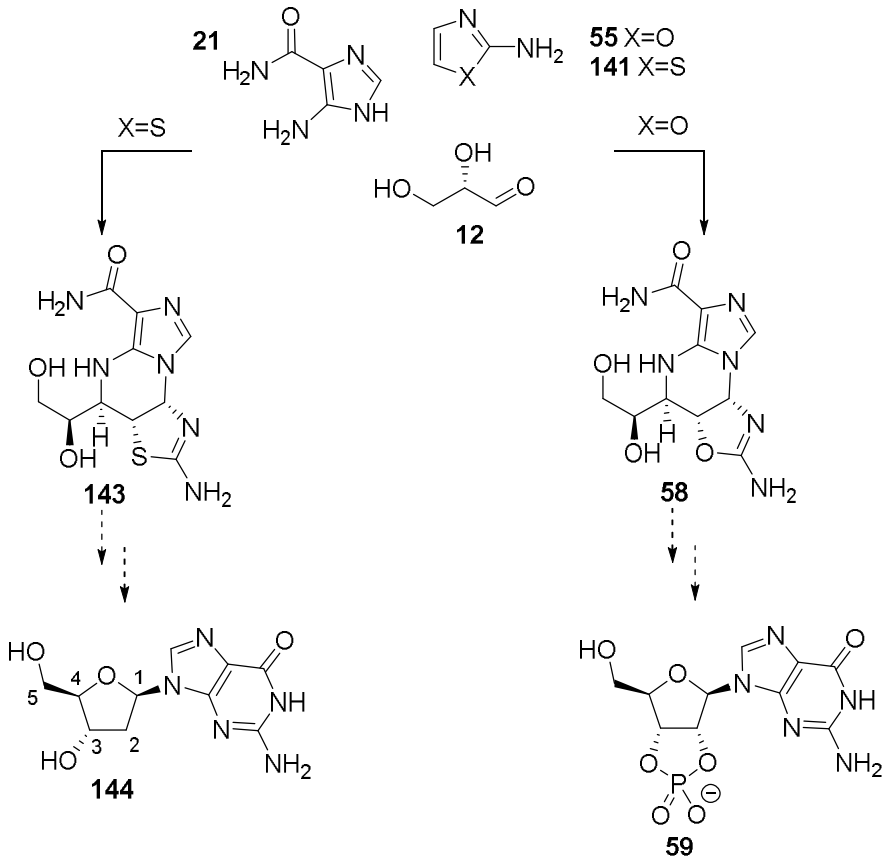
synthesis, linked to the new potential purine synthesis mentioned earlier[274] (*Section 7.1*), uses 2-aminothiazole (**141**). **141** has been shown to form easily in prebiotic conditions (pH 7, in water, at RT) through the rapid reaction of β-mercapto-acetaldehyde (**142**) and cyanamide (**4**) (analogous to reaction to the formation of 2-aminooxazole (**55**) from glycolaldehyde (**3**) and **4**; *Scheme 7.23*)[311]. β-mercapto-acetaldehyde (**142**) is constitutionally similar to glycolaldehyde (**3**), and a compound of interest in other avenues of prebiotic chemistry as a cysteine precursor (*Scheme 7.23*).



*Scheme 7.23: Synthesis of amino-acids serine (**Ser**) and cysteine (**Cys**) from glycolaldehyde (**3**) and β-mercapto-acetaldehyde (**142**) respectively, through Strecker synthesis (top). Synthesis of azoles 2-aminooxazole (**55**) and 2-aminothiazole (**141**) through reaction with cyanamide (**4**) and **3** or **142** respectively. Adapted from Powner et al.*[311]

Powner *et al.* use thiazole **141** in a multicomponent synthesis with AICA **21** and glyceraldehyde (**12**) in order to form a potential DNA precursor: tricyclic product **143** (*Scheme 7.24*), a product analogous to the 3-component product of **21**, 2-aminooxazole (**55**) and **12**: **58** (*Scheme 7.24*). They argue that the lower bond enthalpy of C-S (255 kJ/mol) compared to C-O (355-380 kJ/mol)[311,340] could then lead to a desulphurisation, and in the case of a sulphur positioned at the C2′ of a ribonucleotide (or analogue; which they propose could be attained with product

143) then this desulphurisation could result in DNA monomer deoxyadenosine (**144**) synthesis.

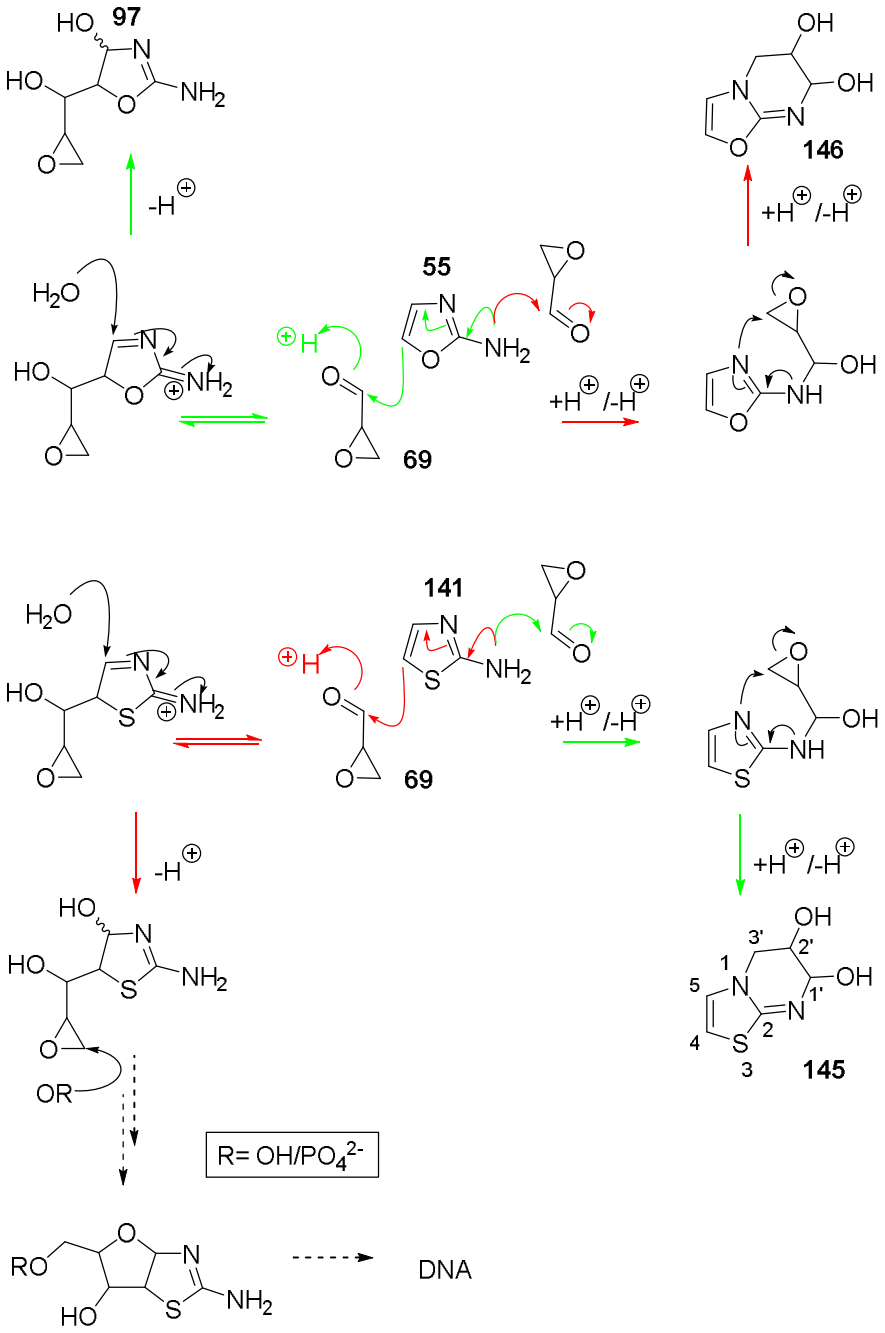


*Scheme 7.24: Divergent synthesis of DNA precursor **143** and purine ribonucleotide precursor **58** leading to deoxyriboadenosine (**144**) and riboadenosine (**59**) respectively, in the multicomponent synthesis of AICA **21**, glyceraldehyde (**12**) and 2-aminooxazole (**55**) or 2-aminothiazole (**141**), as proposed by Powner et al.*[311]

The proposed concomitant route to both ribonucleotides and deoxyribonucleotides could be viewed as problematic by virtue of heterogeneous oligomerisation. While it is argued that the information transfer process, essential to replication, is robust to changes in backbone structure (as seen in the transfer of information from DNA to mRNA for example)[341], the transfer of functional aspects of nucleotides such as RNA, might be more susceptible to changes in the backbone chemistry. However, some nonheritable backbone variety, of the DNA-RNA sort,

has been shown by Trevino *et al.*, to not be "an insurmountable barrier to the emergence of functionality in early nucleic acids"[341] through the study of "mosaic" nucleic acid (mixture of RNA and DNA building blocks) and the successful retention of functionality (adenosine-binding in this case) through rounds of *in vitro* selection.

We proposed to study the reaction of our epoxide glycidaldehyde (**69**) with this prebiotically relevant molecule: 2-aminothiazole (**141**). We hypothesised that 2-amino-tether would likely result in the synthesis of cyclic-hemiaminal thiazole **145** (*Scheme 7.25*), but it was also deemed possible that a masked aldol-type reaction may occur to initially form a species similar to adduct **97** (observed in the reaction of **69** with 2-aminooxazole (**55**)) which could be a precursor to DNA nucleotides (by a similar desulphurisation of the C2′-sulphur as discussed above).



*Scheme 7.25: Comparison of the preferred reaction of glycidaldehyde (**69**) on 2-aminooxazole (**55**) and 2-aminothiazole (**141**): masked-aldol vs 2-amino tether. Top: Reaction of **69** with **55** yielding intermediate **97**. Observed masked-aldol (green) over 2-amino tether (not observed; red). Bottom: Reaction of **69** with **141** yielding cyclic hemi-aminal **145**. Observed 2-amino tether (green) over masked-aldol moiety (not observed; red). Hypothetical route to DNA shown with dotted arrows.*

The reaction of glycidaldehyde (**69**; 140mM) with 2-aminothiazole (**141**; 1 eq) was monitored at RT at pH 5 and 7 in water. Initial results indicated a clean reaction to form what appeared to be thiazole **145** in 81% yield after 11 h (based on ^{1}H NMR analysis; *Scheme 7.25*). Careful analysis of 2D-NMRs confirmed the structure of **145** because of the clear correlation between the C3′ of the sugar moiety and the H-C5 of the aminothiazole (and the reciprocal H-C3′ to C5 correlation) as well as the correlation between C2 and the sugar protons (*Figure 7.16*).

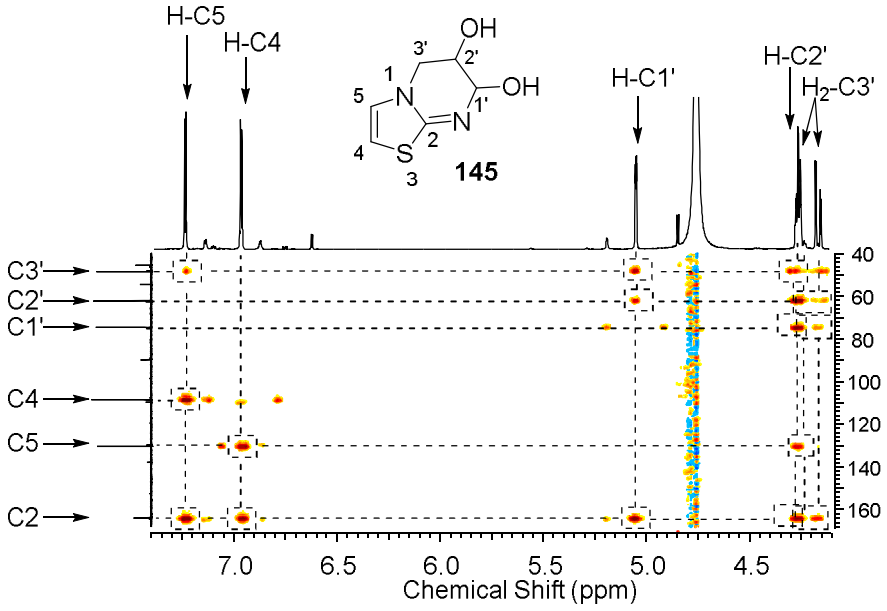


*Figure 7.16: ^{1}H-^{13}C HMBC (600 MHz, H_2O/D_2O 9:1) of the reaction of glycidaldehyde (**69**; 140mM) with 2-aminothiazole (**141**; 1 eq) in H_2O/D_2O (9:1) at pH 7 and at RT. Showing formation of cyclic hemi-aminal **145** (evidenced by the coupling between the H-C5 of the aminothiazole moiety, and the C3' of the sugar moiety and between C2 and H-C1', H-C2' and H-C3'). Only the major anomer of **145** is labelled.*

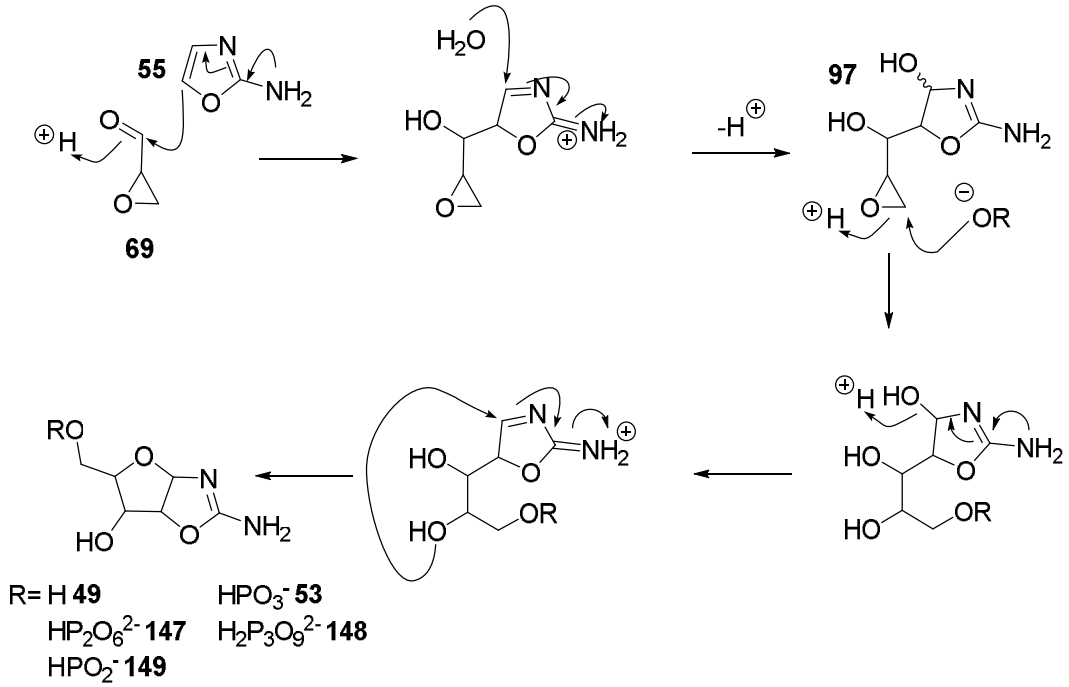
The reaction of **69** and **141** via 2-amino-tether can be explained due to a better energy overlap between the sulfur lone pair and the ring's π-system of 2-aminothiazole (**141**), compared to an oxygen lone pair as seen in 2-aminooxazole (**55**). This increased aromaticity would impede the carbon-carbon bond formation between C1 of **69** and C4 of **141** due to the disfavourable

breaking of aromaticity. A similar explanation was given for the lack of C-C bond formation in the reaction of **141** and glycolaldehyde (**3**) or glyceraldehyde (**12**), instead yielding aminals[311].

Our studies of nucleophilic addition of nucleophiles to glycidaldehyde (**69**) to yield ribopurine (or DNA) precursors has so far been unsuccessful. These results suggested that tethering nucleophiles to the free epoxide was not the right avenue to explore for our purposes. However, the chemistry uncovered gave us better insight into the reactivity of **69** and how it can and cannot be used in prebiotic synthesis.

7.3.2. Phosphorus species

The successful aqueous nucleophilic phosphorylation of oxazoline **97** (the **69**-**55** adduct exploited in our 5′-phosphate aminooxazoline **53** studies detailed in *Section 4*; *Scheme 7.26*) led us to investigate the insertion of other prebiotically relevant nucleophiles. We started by looking at several important phosphorus species such as pyrophosphate (PP_i) and triphosphate (PPP_i), among the highest energy phosphates used by biology[342].



*Scheme 7.26: Proposed mechanism for the reaction of 2-aminooxazole (**55**) with glycidaldehyde (**69**) to form intermediate **97**, and subsequent addition of nucleophiles to form furanosyl-aminooxazolines. The following pairs (nucleophile/product) have been tested/observed experimentally: hydroxide/**49**, P_i/**53**, PP_i/**147**, PPP_i/**148**, HPO_3^-/**149**.*

In order to study the insertion of different phosphorus species in oxazolines **97**, we monitored the reaction epoxide **69** (128mM) and **55** (1.3 eq) at pH 5 until all **69** was consumed and **97** was formed (13 h). We then used separate aliquots and added PP_i or PPP_i and adjusted the solution to pH 7. Upon following the reaction of PP_i addition, ^1H-^{31}P HMBC showed coupling from a diphosphate to the H-C5′ region of aminooxazolines (*Figure 7.17*), confirming the presence of 5′-diphosphate aminooxazolines (**147**) in solution.

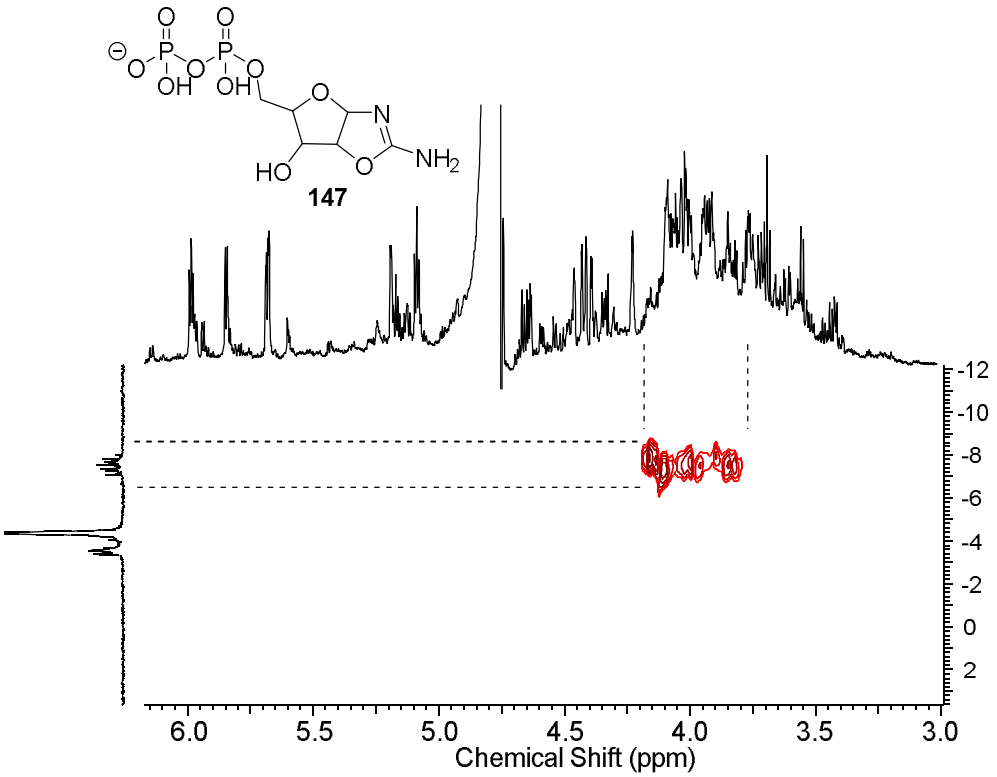


*Figure 7.17: ^{1}H-^{31}P HMBC (400 MHz, H_2O/D_2O 9:1) showing coupling between phosphorus and protons in the 3.6-4.2 ppm region, evidence of successful diphosphorylation at the 5′ position of aminooxazolines. Reaction of gkycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.3 eq) at pH 5, RT for 13 h, followed by addition of pyrophosphate tetrabasic (420mM). Acquisition 5 days after addition of pyrophosphate, at pH 7.*

Similarly, we found that incubation of oxazolines **97** and PPP_i at pH 7 yielded terminal triphosphorylation of the 5′ position of **97**. The characteristic ^{1}H-^{31}P HMBC coupling from a triphosphate to the H-C5′ region of aminooxazolines can be observed (*Figure 7.18*). This evidence strongly suggested formation of 5′-triphosphate aminooxazolines (**148**) in aqueous solution under mild conditions. Further work is needed in order to determine the ratio of **148/147** and non-phosphorylated **49** (as hydrolysis to yield unphosphorylated aminooxazolines (**49**) could be competing with insertion of di/tri-phosphates), as well as the absolute yield of **148/147**. It would also be of interest to determine the diastereomer selectivity in order to corroborate the hypothesis that the diastereomer distribution observed (*ribo>xylo>>lyxo>arabino*) is determined in

the initial **69-55** reaction (as detailed in *Section 4*). These initial results show unprecedented regioselective addition of high energy triphosphate to nucleotides (or its precursors) in prebiotic chemistry and warrants further exploration.

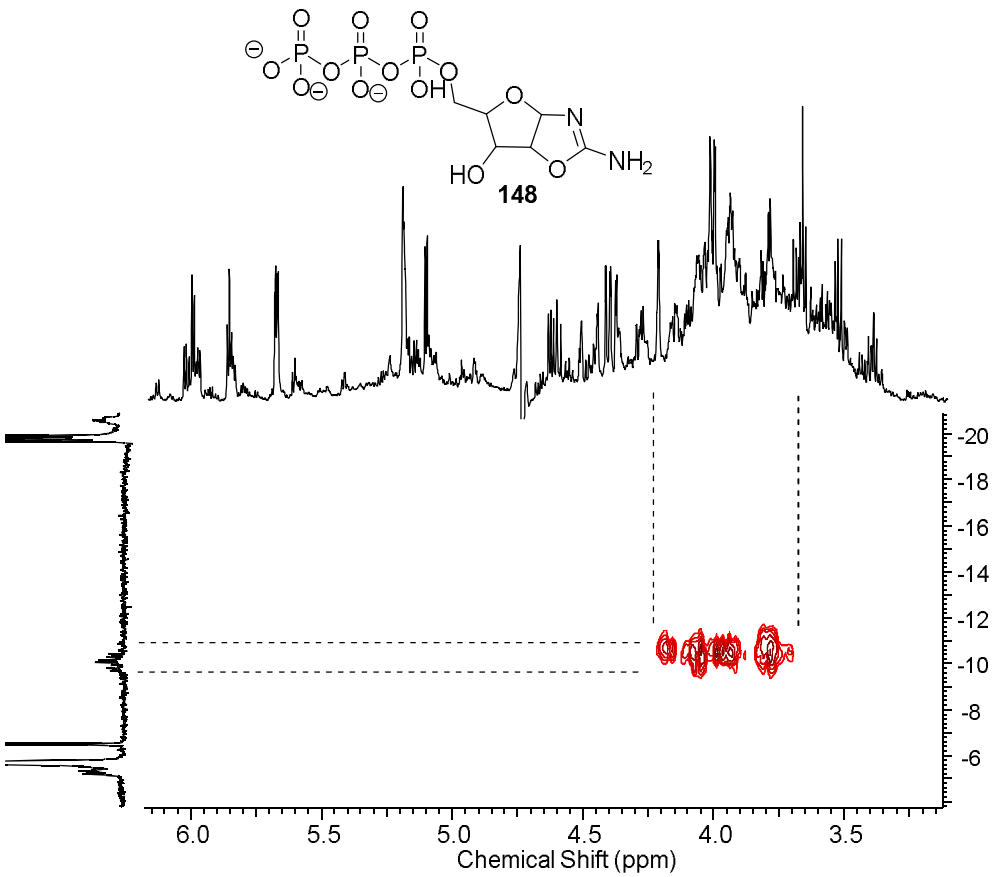
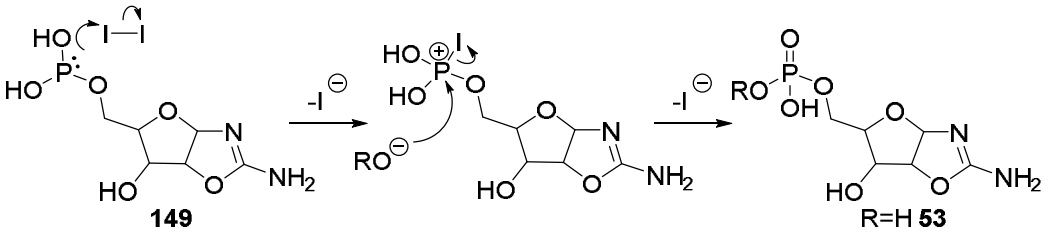


*Figure 7.18: ^{1}H-^{31}P HMBC (400 MHz, H_2O/D_2O 9:1) showing coupling between phosphorus and protons in the 3.7-4.3 ppm region, evidence of successful triphosphorylation at the 5′ position of aminooxazolines. Reaction of glycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.3 eq) at pH 5, RT for 13 h, followed by addition of triphosphate pentabasic (420mM). Acquisition 7 days after addition of triphosphate, at pH 7.*

Another intriguing possibility lay in the nucleophilic phosphitylation. Phosphite H_3PO_3 (oxidation state: +3), the reduced form of phosphate H_3PO_4 (oxidation state: +5), has been argued to have been much more available in prebiotic aqueous environments[343], as concluded in the aqueous corrosion studies of Fe-phosphide and schreibersite by Pasek *et al.*[221,344] as unlike phosphate, phosphite is soluble in saline solutions containing calcium. Phosphite-ester linkages have thus been

suggested as a replacement for phosphate-linkages in an alternative biochemistry preceding the RNA-world[222]. The prebiotic potential of 3′,5′ phosphite linkages has also been studied, and it was concluded that these types of linkages were too unstable to have played a long-lasting role, unless they were "rapidly converted to a more stable form perhaps by oxidation to phosphate compounds."[345] A 5′-phosphite aminooxazoline (**149**) could lead to 5′-phosphite nucleotides (through similar development of the nucleobase moiety) which then opens up the possibility of oxidative nucleotide linkage (by attack of a 3′-OH from another ribonucleotide on the activated phosphite; *Scheme 7.29*, where R is another ribonucleotide).



*Scheme 7.29: Proposed mechanism for the oxidation of 5′-phosphite aminooxazoline (**149**) to 5′-phosphate aminooxazoline (**53**) using model oxidant I_2.*

Similarly to the reactions described above, we studied the incubation of oxazolines **97** and phosphite (420mM) over 5 days at pH 7, and RT. Results showed clear ^{1}H-^{31}P HMBC correlation between a phosphite and the H-C5′ region of aminooxazolines (*Figure 7.19*), suggesting successful insertion of phosphite at the 5′ position of **97**. There is precedent for the prebiotic oxidation of phosphite[220] and so we were encouraged to assess the feasibility of oxidation of the phosphite incorporated to our 5′-phosphite aminooxazoline (**149**). After 11 days, iodine I_2 (420mM) was added to the reaction as a model oxidant. Due to excess phosphite used in the initial reaction (420mM *vs* 128mM of **69** initially used), the reaction proceeded sluggishly. After 20 days, we added another equivalent of I_2 in order to push the reaction further. Another 13 days at RT showed successful oxidation of the 5′-phosphite to yield 5′-phosphate aminooxazolines (**53**; determined by ^{1}H-^{31}P

HMBC analysis, *Figure 7.20*). This proof of concept experiment demonstrated the possibility of oxidation of a tethered phosphite. Excitingly, this opens up the possibility of oxidative oligomerisation (which we anticipate will be explored further in the future).

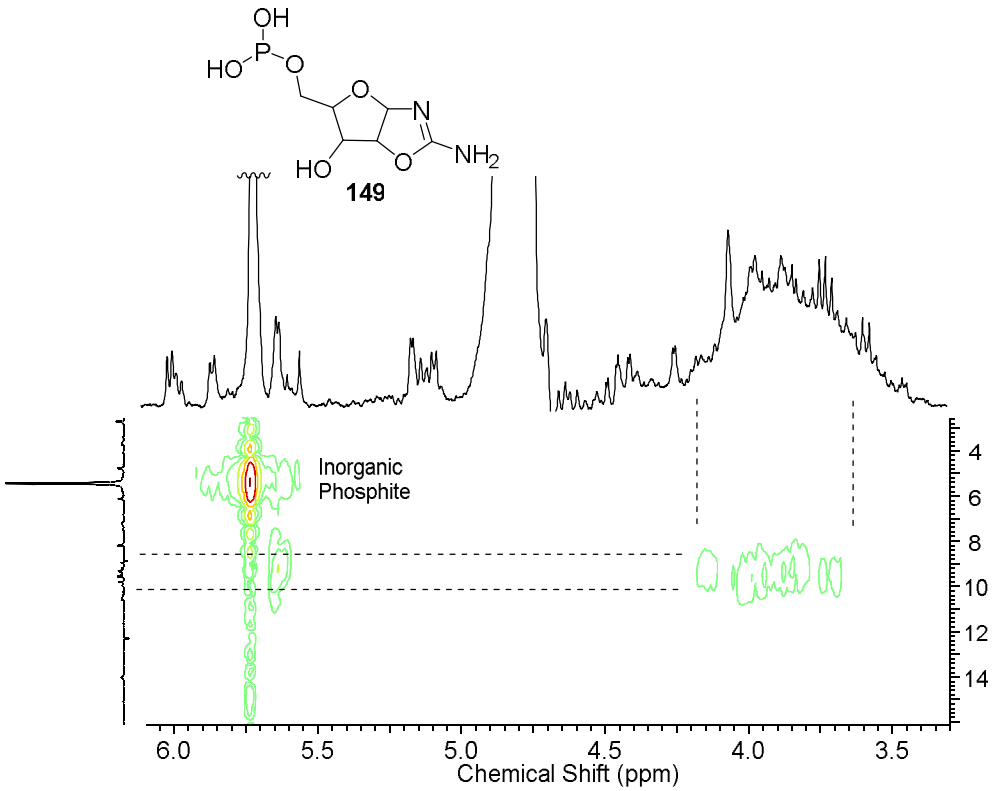


*Figure 7.19: ^{1}H-^{31}P HMBC (400 MHz, H_2O/D_2O 9:1) showing coupling between phosphorus and protons in the 3.6-4.2 ppm region, evidence of successful phosphite insertion at the 5' position of aminooxazolines. Reaction of gkycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.3 eq) at pH 5, RT for 13 h, followed by addition of phosphite dibasic (420mM), at pH 7. NMR acquisition 5 days after addition of phosphite.*

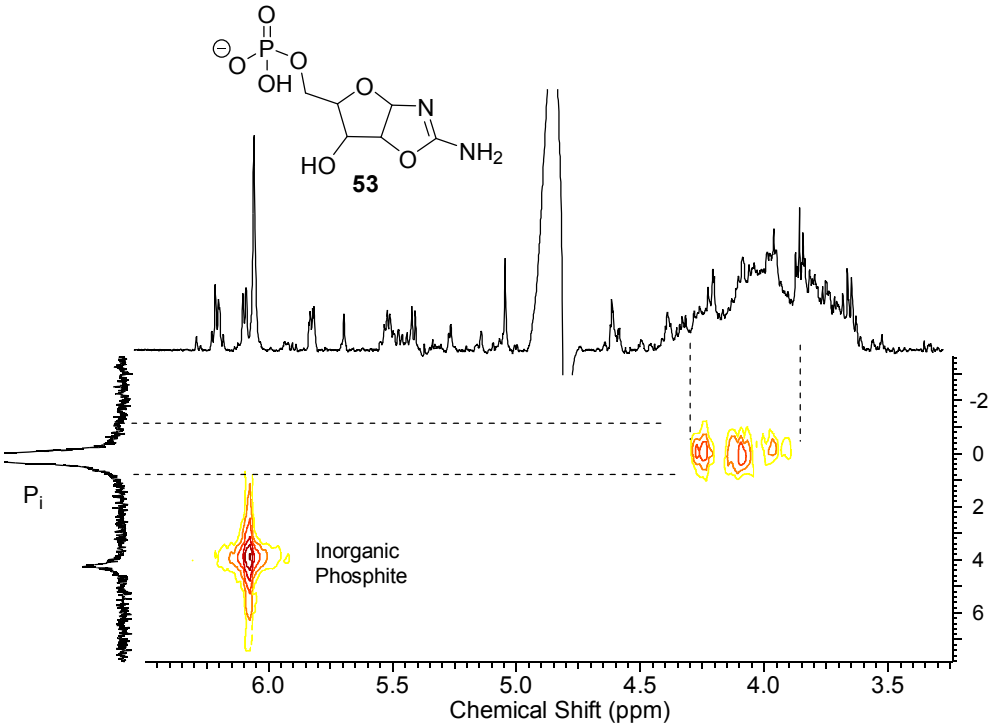


*Figure 7.20: ^{1}H-^{31}P HMBC (400 MHz, H_2O/D_2O 9:1) showing coupling between phosphorus and protons in the 3.6-4.2 ppm region, evidence of successful phosphite oxidation at the 5′ position of aminooxazolines. Reaction of gkycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; 1.3 eq) at pH 5, RT for 13 h, followed by addition of phosphite dibasic (420mM). Reaction was carried out at pH 7 for 11 days, at which point iodine (420mM) was added. After 20 days, iodine (420mM) was added again. NMR acquisition 13 days after second addition of iodine.*

These nucleophilic insertions of various phosphorus species demonstrate the robustness of the nucleophilic 5′-substitution in **97**. Not only have we shown a novel way of prebiotically inserting an activated phosphate at the canonical 5′-position, we also opened up the door to another avenue not yet explored in prebiotic chemistry: oxidative linkage.

Furthermore, this chemistry solves issues experienced earlier, whereby the aldehydic component of glycidaldehyde (**69**) reacts easily with nitrogenous purine precursors. One could imagine that using a tethered version of **69**, such as oxazolines **97** (where the aldehyde initially reacted without a subsequent cyclisation onto the epoxide moiety), opens up the possibility of nucleophilically

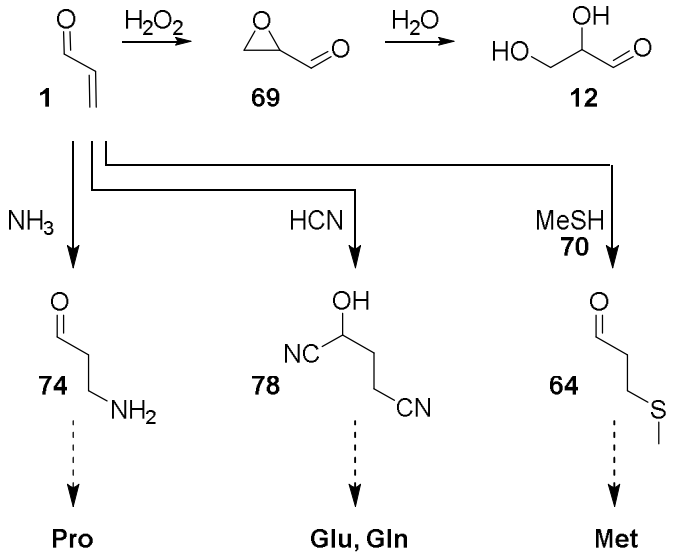
inserting a purine precursor at the 5′-position of aminooxazolines. Using **97** would bypass a C3-modified glyceraldehyde and bring us straight to the desired key intermediates mentioned in this section (*Scheme 7.6*, *vide supra*)[xxxiv].

[xxxiv] These reactions are currently being investigated in the Powner lab.

8. Conclusion

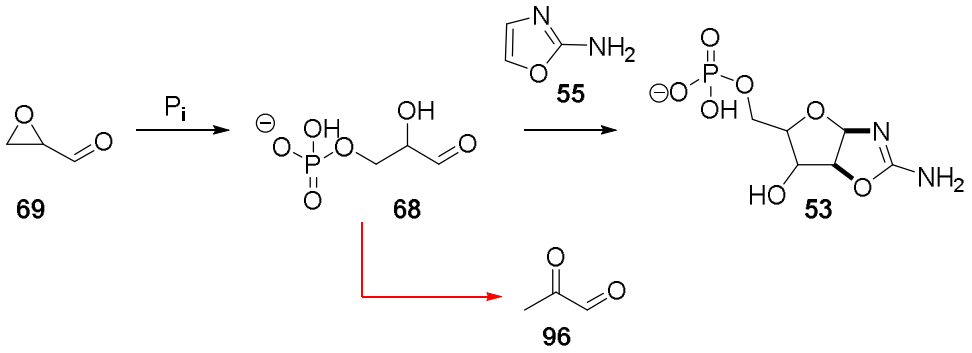
During our research into prebiotic chemistry, new ways of connecting different metabolite classes (amino acids and nucleotides) have been uncovered, as well as new avenues that could be explored in future prebiotic research. Excitingly, the first prebiotically plausible nucleophilic and selective aqueous phosphorylation was established, by-passing the need for dry-state phosphorylation.

Acrolein (**1**) has been demonstrated to be both an amino acid precursor, through Michael addition at the C3 position with cyanide, ammonia or methanethiol (**70**) leading to Strecker-aldehydes of glutamate/glutamine, proline and methionine, respectively (*Scheme 8.1*); and a precursor to glyceraldehyde (**12**), a known nucleotide building block, through epoxidation to glycidaldehyde (**69**) and subsequent hydrolysis (*Scheme 8.1*). Strikingly, the successful terminal derivatisation in amino acids using **1** hinted at a possible use of **1** in terminal phosphorylation of precursors *en route* to ribonucleotides 5′-phosphate.



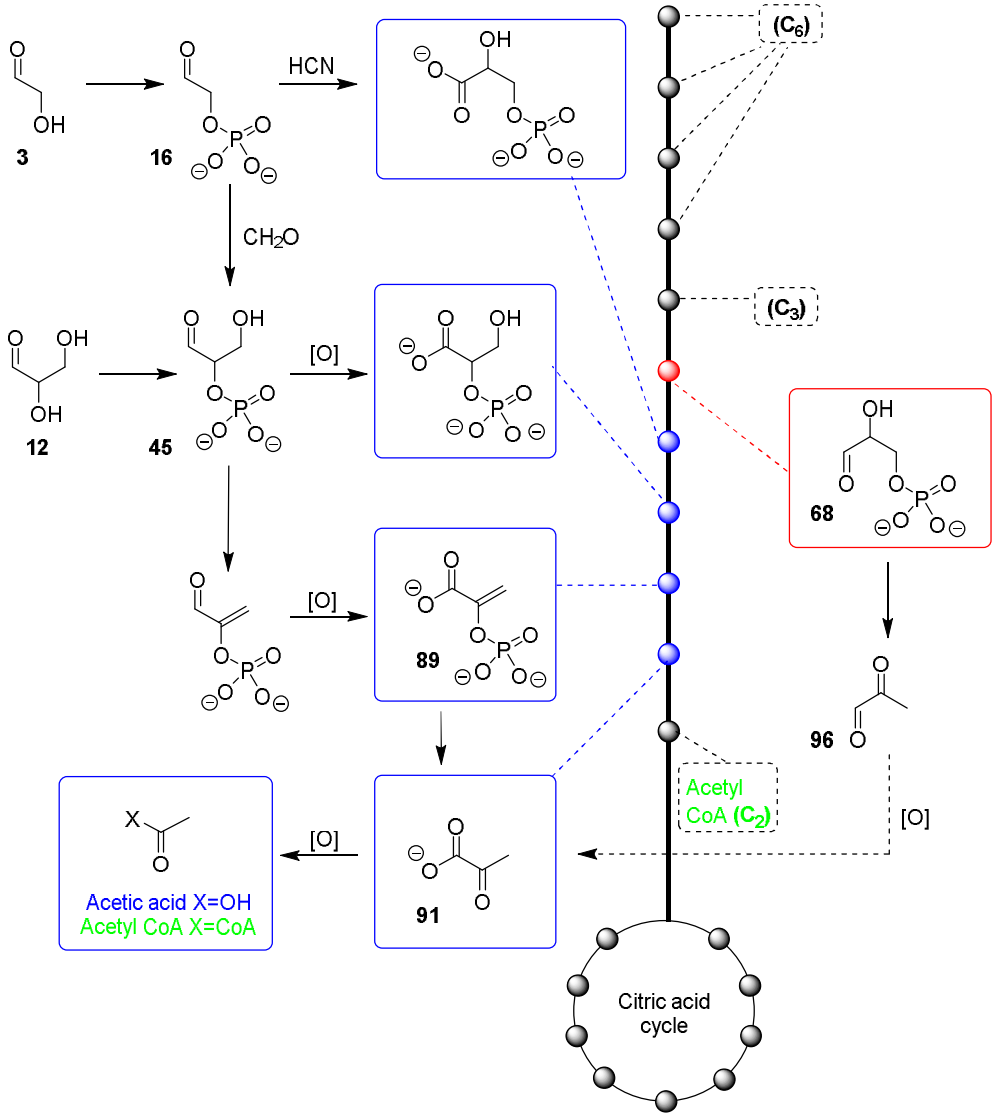
*Scheme 8.1: Michael addition of nucleophiles ammonia, cyanide, methanethiol (**70**) and hydrogen peroxide to acrolein (**1**) resulting in the synthesis of precursors to amino acids proline (**Pro**), glutamine/glutamate (**Gln**/**Glu**), and methionine (**Met**), as well as nucleotide precursor glyceraldehyde (**12**), respectively.*

The epoxidation of **1** with the prebiotically plausible simple oxidant H_2O_2 allowed for access to new reactivity compared to glyceraldehyde (**12**) in the epoxide moiety of **69**. Incubation of epoxide **69** with phosphate allowed for transient access to glyceraldehyde-3-phosphate (**68**), followed by rapid E1cB elimination to methylglyoxal (**96**). Interestingly, **68**, a metabolite in extant biology, had also been linked to nucleotide synthesis through reaction with 2-aminooxazole (**55**) to form 5′-phosphorylated aminooxazolines (**53**)[283] (*Scheme 8.2*).



*Scheme 8.2: Phosphorylation of glycidaldehyde (**69**) to glyceraldehyde-3-phosphate (**68**). Reaction of **68** and 2-aminooxazole (**55**) leads to nucleotide precursor 5′-phosphorylated aminooxazoline (**53**). However, **68** undergoes rapid E1cB phosphate elimination to form methylglyoxal (**96**; red pathway).*

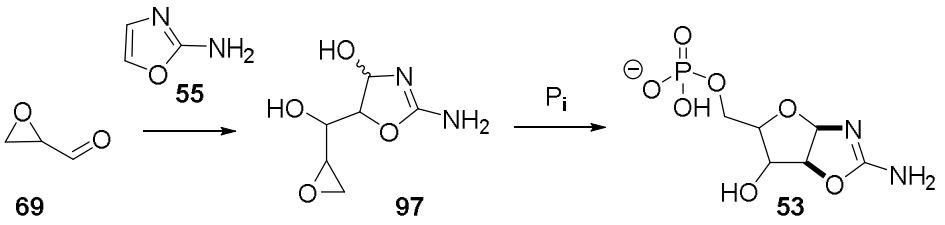
It is noteworthy that the degradation product of **68**, methylglyoxal (**96**) is a potential precursor to essential metabolite pyruvate (**91**) through aldehyde-oxidation. Recent oxidation studies of metabolites in the Powner lab were met with success and provided a link between nucleotide precursors such as glycolaldehyde (**3**) and glyceraldehyde (**12**) with highly conserved essential metabolites of the glycolysis pathway[290]. We propose that an additional link between nucleotide synthons and metabolism could be made through **68** and its degradation pathway if a prebiotically plausible oxidation of **96** to pyruvate (**91**) can be found (*Scheme 8.3*).



*Scheme 8.3: Prebiotic link between nucleotide precursors glycolaldehyde (**3**) and glyceraldehyde (**12**) and the highly conserved metabolic network (blue) leading to the citric acid cycle through phosphorylation and oxidation, as reported by Coggins and Powner[290]. Proposed link between metabolite glyceraldehyde-3-phosphate (**68**; red) and pyruvate (**91**) through phosphate elimination to methylglyoxal (**96**) and hypothetical oxidation.*

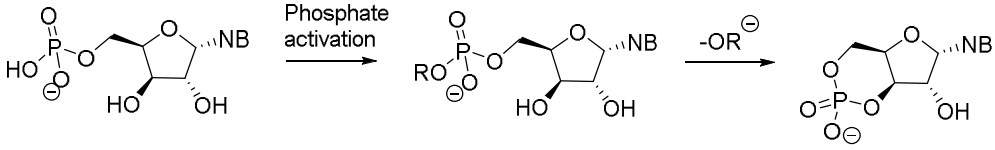
In the context of nucleotide synthesis, we found that reversing the reaction order in the synthesis of 5′-phosphorylated aminooxazolines (**53**), by reacting epoxide **69** with 2-aminooxazole (**55**) initially (to form adduct **97**) and then incorporating phosphate, blocked the elimination pathway and led to the regiochemically controlled terminal phosphorylation of **53** (*Scheme 8.4*). This synthesis provides

the first regioselective 5′-phosphorylation of ribonucleotide precursors from prebiotically plausible precursors. Moreover, it completely bypasses the need for a change of reaction conditions, such as the dry-state phosphorylations used in previous ribonucleotide syntheses[33,225], or the need for (low-yielding) phosphate activation in the phosphorylation step.



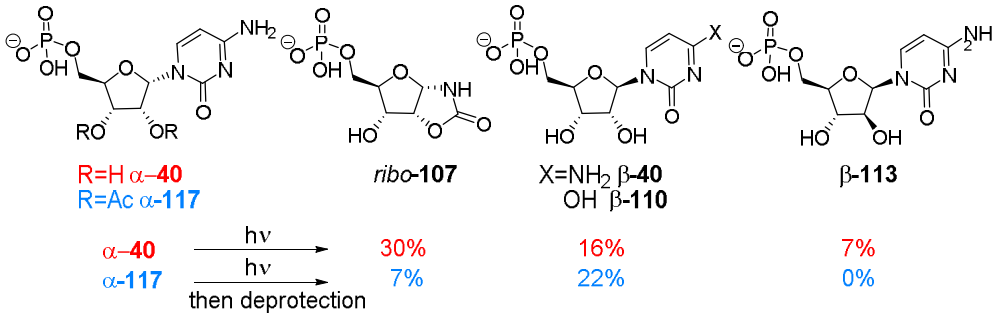
*Scheme 8.4: New prebiotic nucleophilic phosphorylation involving the synthesis of adduct **97** from glycidaldehyde (**69**) and 2-aminooxazole (**55**) followed by regiospecific phosphorylation yielding nucleotide precursor 5′-phosphorylated aminooxazoline (**53**).*

The high *ribo*-selectivity in **53** synthesis is further amplified by the facile interconversion of *arabino*-**53** to *ribo*-**53**. Notably, *lyxo*-**53** was not observed in the synthesis of **53**, resulting in a simple reaction mixture composed of only two species, *ribo*-**53** and *xylo*-**53**. Both *ribo*-/*xylo*-**53** form their respective furanosyl 2,2′-anhydrocytidine 5′-phosphate (**112**) quantitatively though reaction with cyanoacetylene (**6**), which are then easily hydrolysed to the corresponding nucleotides. The *cis*- disposition of the C3′ and C4′ positions in *xylo*- compounds means that in order to oligomerise xylonucleic acids, a synthetic protecting group strategy needs to be employed in order to block cyclisation of activated phosphate and the proximal C3′-OH[32,346,347]. In the context of prebiotic chemistry, lengthy protecting group strategies are not plausible and therefore xylonucleic acids can be considered dead-end monomers (*Scheme 8.5*).



Scheme 8.5: Phosphate activation of xylonucleic acids results in intramolecular attack of the C3′-hydroxyl, resulting in dead-end monomers. NB = nucleobase.

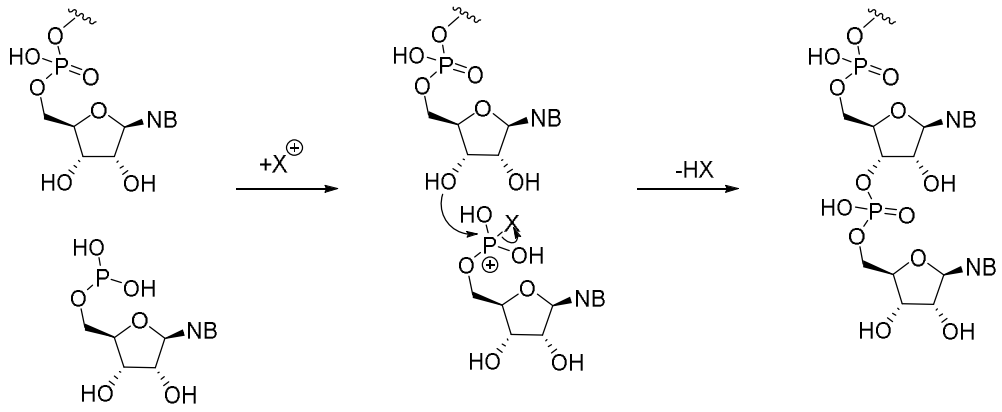
The synthesis described thus far leaves us with one compound of interest: α-ribocytidine 5′-phosphate (α-**40**), which is only one stereochemical inversion away from the canonical β-nucleotide β-**40**. UV irradiation of α-**40** leads to 16% anomerisation of C1′ to yield canonical β-**40**[323]. Although promising, the reaction is accompanied by a plethora of unwanted chemistry including notably C2′-epimerisation to arabinonucleotides **113** and formation of oxazolidinone 5′-phosphate (**107**), destroying the nucleobase moiety. The complexity of the products obtained clouded the benefits of the anomerisation. A solution was found in the temporary blockade of the C2′ hydroxyl by prebiotic acetylation to yield 2',3'-(diacetyl)-α-ribocytidine-5'-phosphate (α-**117**), which led to increased C1′-anomerisation under UV irradiation (22%). Acetylated compound α-**117** is easily prebiotically deprotected by incubation with ammonia. Remarkably, acetylation of the C2′-position suppressed deleterious C2′-epimerisation and limited oxazolidinone **107** formation (*Scheme 8.6*).



*Scheme 8.6: Results of the irradiation of 5'-phosphate α-ribocytidine (α-**40**; red) showing significant base destruction to oxazolidinones 5'-phosphate (**107**) and C2'-epimerisation to 5'-phosphate β-arabinocytidine (β-**113**); alongside improved selectivity through irradiation of 2',3'-O-diacetyl-α-ribocytidine-5'-phosphate (α-**117**; blue) showing limited **107** formation and no C2'-epimerisation. Results of α-**117** are reported after de-acetylation.*

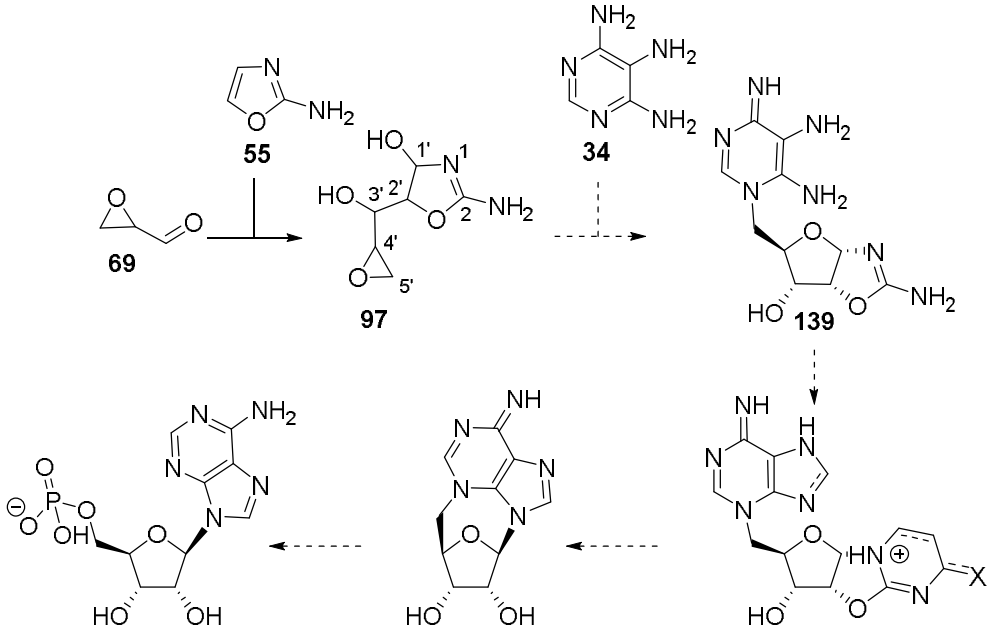
The completed studies demonstrate the first plausible regiospecific synthesis of nucleotide-5'-phosphates in water, starting with simple prebiotic feedstock molecules. Furthermore, adduct **97**, through the retention of glycidaldehyde's epoxide moiety, can be used to regiospecifically introduce other nucleophiles. Initial results suggested successful incorporation of polyphosphates, which would

lead to the first prebiotic synthesis of naturally activated ribonucleotide monomers in water. Interestingly, adduct **97** can also be used to introduce other phosphorus species, such as phosphite. Phosphite has the advantage of being more soluble than phosphate in the presence of divalent metal ions, and as such could have been prebiotically more available than phosphate in aqueous solution[217]. Monomers containing a 5′-phosphite open up a new intriguing avenue for oligomerisation through oxidative ligation (*Scheme 8.7*).



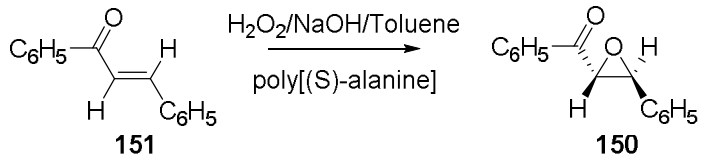
Scheme 8.7: Theoretical use of 5′-phosphite ribonucleotides in the oligomerisation process through oxidative linkage. NB = nucleobase.

Adduct **97** also opens up the avenue for other nucleophilic 5′-substitutions, such as purine precursors 4,5,6-triaminopyrimidine (**34**) or AICA **21**. Nucleophilic insertion of these purine precursors would lead to a 5′-modified aminooxazoline, which through additional steps (activation towards glycone exchanges, purine elaboration and 5′-displacement, as discussed in detail in *Section 7*), could lead to a prebiotic and regioselective synthesis of purine nucleotides (*Scheme 8.8*).



*Scheme 8.8: Hypothetical route from adduct **97**, accessed through glycidaldehyde (**69**) and 2-aminooxazole (**55**), to purine ribonucleotides by regiospecific insertion of a purine precursor at the C5′ of **97**, as exemplified with 4,5,6 triaminopyrimidine (**34**).*

Discovering the new use of prebiotic building blocks opens up new opportunities for addressing the chirality in these compounds. The centrality of epoxide **69** to the synthesis of nucleotide-5′-phosphates, as seen in the research described throughout, opens new opportunities for addressing its chiral synthesis at the origin of life. There is precedent for the chiral synthesis of epoxides, that although not prebiotic in themselves, could potentially be translated to prebiotic research. Juliá *et al*. describe the 90% *ee* synthesis of epoxide **150** by using homochiral polypeptides such as polyalanine to stereoselectivity direct the epoxidation of chalcone (**151**) by H_2O_2[348] (*Scheme 8.9*).



*Scheme 8.9: Epoxidation of chalcone (**151**) by H_2O_2 in the presence of homochiral poly[(S)-alanine], resulting in a 90%* ee *in the epoxide **150**, as reported by Juliá et al.*[348]

The homochirality of amino acids has been studied extensively ever since it was found that meteorites contain a significant *ee* in favour of the canonical L-amino acids (up to L_{ee}=18.5% for L-isovaline in the Murchison meteorite[127]). This *ee* can be amplified for example by selective crystallisation of racemic crystals, thereby increasing *ee* in solution (from 1% to 90% as seen in the studies by Breslow and Levine[132]; or Klussmann *et al.*[133]; *vide supra*, *Section 1.1.3.3*). The possibility of a chiral synthesis of epoxides by exploitation of polypeptide catalysis is intriguing in the context of prebiotic research as we are always looking to improve the connection between different classes of biologically relevant molecules (*Section 1.3.2*).

It is important to note that the results described throughout showcase the products obtained (from wanted and un-wanted reactions) to the best of our ability. However, we are aware that in some instances it was not possible to describe the complexity of reaction mixtures in their entirety. As such, internal standards were used to describe yields when possible. With that in mind, the selectivity of the reactions must be interpreted with full knowledge of these limitations.

Our aim in the field of prebiotic chemistry is to give a comprehensive overview of the chemistry at play in the origin of life. During the research into the synthesis of the possible first biomolecules, it is important to connect all the relevant results obtained with other aspects of life such as compartmentalisation and metabolism, in the hopes of gaining insight into the *overarching* processes that may have occurred in the early steps of biochemical evolution. Some important new connections between different metabolite classes have been highlighted during these studies bringing us closer to understanding possible avenues life may have taken during its inception. Furthermore, we have demonstrated simplified new routes to ribonucleotides, solving some problematic steps encountered in previous prebiotic research. There is still a long way to go in answering the questions that have been asked by the scientific community over thousands of years: 'What is life?' and 'How did life originate?', but we hope that through the research described here we have taken a small but significant step towards one day understanding the processes that governed the chemical evolution of life.

9. Experimental

9.1. General

Reactions requiring anhydrous conditions were carried out under an argon or nitrogen atmosphere, using oven-dried glassware. Reagents and solvents obtained from commercial sources and used without further purification. Deionised water was obtained from an *Elga Option 3* purification system. Dowex® 50W×8 resin was purchased from *Acros Organics* and regenerated with HCl solution. Solution pH values were measured using a Mettler Toledo Seven Compact pH meter with a Mettler Toledo InLab semi-micro pH probe. The readings for D_2O solutions are reported as pD. The readings for H_2O and H_2O/D_2O solutions are reported uncorrected. ^1H NMR spectra (H_2O/D_2O) are solvent suppressed (noesygppr1d) with presaturation and spoil gradients.

TLC analysis was on Merck Kieselgel 60 F_{254} coated aluminium sheets, with UV visualisation at 254 nm, or visualisation by staining with $KMnO_4$, vanillin, cerium ammonium molybdate or 10% H_2SO_4 in EtOH. Flash column chromatography was carried out using Merck silica gel 60 (40-60 μm). Automated flash column chromatography was carried out using *Biotage Isolera Four* purification system and Telos Silica Flash-LL columns.

^1H and ^{13}C NMR spectra were recorded using Avance III 600 Cryo (600 MHz), and are reported as follows: chemical shift (multiplicity, coupling constants (*J*), number of protons, nuclear assignment) with δ in ppm, *J* in Hz, and assignments by COSY, HMBC, DEPT, HSQC. ^{31}P NMR spectra were recorded using AVANCE 300 equipped with a gradient probe (300 MHz) or Bruker Avance III equipped with a gradient probe (400 MHz). Spin multiplicities are indicated by the following symbols: s (singlet); d (doublet); t (triplet); q (quartet); qn (quintet); m (multiplet); ABX geminal (AB) spin system coupled to one other nuclei (X)); ABXY (germinal (AB) spin system coupled to two other nuclei (X, Y)); or a combination of these. ^1H and ^{13}C chemical shifts relative to TMS were calibrated

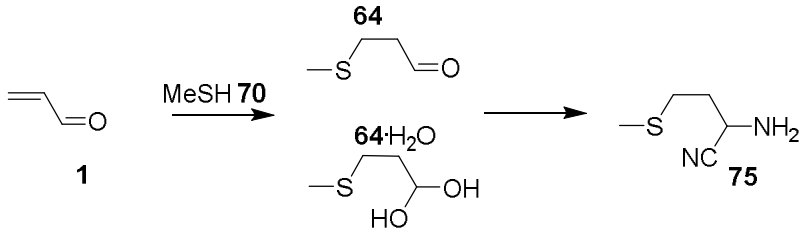
using the residual solvent peak. Unless otherwise specified, spectra were recorded at 298 K. When numbers of compounds in solution over time are given, the numbers represent the percentage of each compound in solution based on NMR integration compared to total compounds in solution, unless otherwise stated. Melting points were determined using a Reichert Hotstage apparatus for all solids where possible and are quoted to the nearest ºC and are uncorrected. Infrared spectra (IR) were recorded on a Perkin Elmer Spectrum 100 FT-IR spectrometer. Absorption maxima are reported in wavenumbers (cm^{-1}). Mass spectra and accurate mass measurements were recorded on a VG70-SE or MAT 900X instrument at the University College London Chemistry Department. Optical rotation ($[\alpha]_D^{25}$) were recorded with a Perkin Elmer polarimeter with sodium light (l= 589 nm). Values are given as a mean of at least 5 measurements using the following equation: $[\alpha]_D^{25}=100\alpha/lc$ (α= observed angle of rotation; l= cell length in dm (1 dm); c=concentration of sample in g/100 mL; measured at 25°C).

9.2. Experimental Procedure for 2. The use of acrolein in prebiotic synthesis

General procedure for Strecker synthesis

To a solution of aldehyde (0.5 mL, 0.14mM, 70 μmol) was added potassium cyanide (6.8 mg, 105 μmol), followed by ammonia (48.8 μL of 30% aqueous NH_4OH solution, 350 μmol). The solution was adjusted to pH 9.5 and the reaction monitored by ^1H NMR over time.

*Formation of 3-(methylthio)propanal (**64**) from acrolein (**1**) with methanethiol (**70**)[288], followed by Strecker synthesis[289]*



Methanethiol (**70**)[xxxv] was bubbled continuously through a solution of P_i (0.5 mL, 1M, pH 7), whilst a solution of acrolein (**1**; 9.3 μL, 0.14 mmol, 0.5 mL in H_2O/D_2O 9:1) was added dropwise over 5 min, yielding 3-(methylthio)propanal (**64**; >94% yield, as a mixture of aldehyde **64** and its hydrate **64**·H_2O), confirmed by spiking with an authentic sample. General Strecker procedure was followed on **64** at pH 10, yielding 2-amino-4-(methylthio)butanenitrile (**75**) after 20 h at RT.

[xxxv] **70** was generated by dropping sodium methanethiolate (21% aqueous solution) onto mono-basic sodium phosphate.

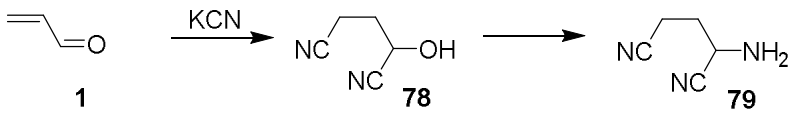
^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for the pH described for a given reaction)

3-(Methylthio)propanal hydrate (**64**·H_2O): 5.07 (t, *J* = 5.5 Hz, 1 H, H-(C1)); 2.51 (t, *J* = 7.6 Hz, 2 H, H-(C3)); 2.04 (s, 3 H, SCH_3); 1.81 (dt, *J* = 7.6, 5.5 Hz, 2 H, H-(C2)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 90.6 (C1); 37.2 (C2); 29.2 (C3); 15.0 (SCH_3).

3-(Methylthio)propanal (**64**): 9.60 (s, 1 H, H-(C1)); 2.80 (t, *J* = 6.6 Hz, 2 H, H-(C3)); 2.73 (t, *J* = 6.6 Hz, 2 H, H-(C3)); 2.04 (s, 3 H, SCH_3[xxxvi]). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 207.4 (C1); 43.0 (C3); 26.2 (C2); 14.9 (SCH_3).

2-Amino-4-(methylthio)butanenitrile (**75**): 3.96 (t, *J* = 7.2 Hz, 1 H, H-(C1)); 2.69-2.54 (m, 2 H, H-(C3)); 2.06 (s, 3 H, SCH_3); 1.99 (m, 2 H, H-(C2)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 123.1 (CN); 42.4 (C1); 33.8 (C2); 29.6 (C3); 14.8 (SCH_3).

*Formation of 2-hydroxypentanedinitrile (**78**) from acrolein (**1**) with cyanide*[37]*, followed by Strecker synthesis*[289]



1 **78** **79**

Potassium cyanide (45.6 mg, 0.7 mmol) was dissolved in H_2O/D_2O (9:1, 1 mL) and the solution was adjusted to pH 9.5 using 1M HCl. Acrolein (**1**; 9.3 µL, 0.14 mmol) was added, the solution was adjusted to pH 9 with 1M HCl. The reaction was then incubated at RT and ^{1}H NMR spectra were periodically acquired. showing conversion to 2-hydroxypentanedinitrile (**78**; 65% yield by ^{1}H NMR integration, Appendix - *Figure 11.5*, *a*.) after 24 h. General Strecker procedure was followed (omitting the addition of potassium cyanide) on **78** at pH 10.5, yielding 2-aminopentanedinitrile (**79**) after 7 days at RT.

[xxxvi] This peak is under the **64**·H_2O S$\underline{CH_3}$ peak.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for the pH described for a given reaction)

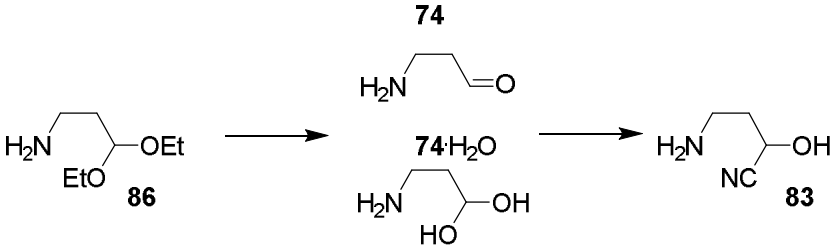
2-Hydroxypentanedinitrile (**78**): 4.75 (H-(C1), *obscured by the residual solvent peak*); 2.68 (t, *J* = 7.1 Hz, 2 H, H-(C2)); 2.27-2.08 (m, 2 H, H-(C3)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 120.8 (CN); 60.0 (C1); 30.6 (C3); 13.3 (C2).

2-Aminopentanedinitrile (**79**): 4.11 (t, *J* = 7.2 Hz, 1 H, H-(C1)); 2.80 (td, *J* = 7.3, 1.9 Hz, 2 H, H-(C3)); 2.31-2.17 (m, 2 H, H-(C2)).

^{1}H NMR (600 MHz, D_2O) δ ppm (shifts are pH dependent, given here for the pD 10)

2-Hydroxypentanedinitrile (**78**): 4.75 (H-(C1), *obscured by the residual solvent peak*); 2.66 (d, *J* = 7.2 Hz, 2 H, H-(C2)), 2.19 (m, 2 H, H-(C2)). ^{13}C NMR (151 MHz, D_2O) δ_C 121.0 (CN); 60.0 (C1); 30.5 (C3); 13.2 (C2).

*Synthetic standard of 3-aminopropanal (**74**), followed by cyanohydrin formation*



3-Aminopropanal (**74**) was prepared following a procedure adapted from Patel *et al.*'s synthetic **74** synthesis[37]. A solution of 1-amino-3,3-diethoxypropane (**86**; 0.48 mL, 3 mmol) in ice-cold aqueous HCl (3 mL, 3M) was prepared and warmed to RT over 1 h. The solution was stirred for an additional 2 h at which point it was diluted with ice-cold H_2O (10 mL), and analysed by ^{1}H NMR, showing quantitative conversion to 3-aminopropanal (**74**; in the form of its hydrate **74**·H_2O). Potassium cyanide (14.45 mg, 0.22 mmol) was added to an aliquot (1 mL) of the solution and the resulting solution was adjusted to pH 3 using 4M

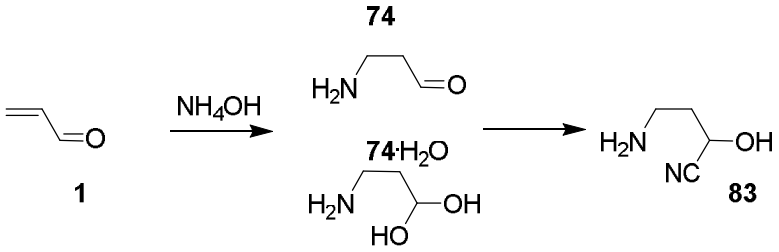
NaOH. ^{1}H NMR spectra were periodically acquired, showing formation of 4-amino-2-hydroxybutanenitrile (**83**) after 6 h.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for the pH described for a given reaction)

3-Aminopropanal hydrate (**74**·H_2O): 5.15 (t, *J* = 5.9 Hz, 1 H, H-(C1)); 3.05 (t, *J* = 6.4 Hz, 2 H, H-(C3)); 1.89 (dt, *J* = 6.4, 5.9 Hz, 2 H, H-(C2)).

4-Amino-2-hydroxybutanenitrile (**83**): δ 4.85 (dd, *J* = 7.6, 5.1 Hz, 1 H, H-(C1)); 3.25 (t, *J* = 7.3 Hz, 2 H, H-(C3); 2.32-2.17 (m, 2H, H-(C2)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 123.1 (CN); 58.4 (C1); 37.0 (C3); 32.4 (C2).

*Formation of 3-aminopropanal (**74**) from acrolein (**1**) and ammonia, followed by cyanohydrin formation*[37]



Ammonium chloride (37.5 mg, 0.7 mmol) was dissolved in H_2O/D_2O (9:1, 1 mL) and the solution was adjusted to pH 6.5 using 1M NaOH. Acrolein (**1**; 9.3 μL, 0.14 mmol) was added, the solution adjusted to pH 6.5 and ^{1}H NMR spectra were periodically acquired. The solution was re-adjusted to pH 6.5 after 24 h (Appendix - *Figure 11.6*, *a*.). Potassium cyanide (6.8 mg, 105 μmol) was added to an aliquot (0.5 mL) of the solution and the solution was adjusted to pH 3 and ^{1}H NMR spectra were periodically acquired, showing formation of 4-amino-2-hydroxybutanenitrile (**83**) after 6 h.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for the pH described for a given reaction)

3-Aminopropanal hydrate (**74**·H_2O): 5.21 (t, *J* = 5.4 Hz, 1 H, H-(C1)), 3.12 (t, *J* = 7.0 Hz, 2 H, H-(C3)), 1.95 (dt, *J* = 7.0, 5.4 Hz, 2 H, H-(C2)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 89.7 (C1); 37.6 (C3); 34.7 (C2).

3-Aminopropanal dimer (**85**): 5.16 (t, *J* = 5.6 Hz, 1 H, H-(C1)), 3.63 (t, *J* = 6.8 Hz, 2 H, H-(C3)), 1.84 (dt, *J* = 6.8, 5.6 Hz, 2 H, H-(C2)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 89.7 (C1); 58.8 (C3); 40.6 (C2).

4-Amino-2-hydroxybutanenitrile (**83**): δ 4.84 (dd, *J* = 7.6, 5.1 Hz, 1 H, H-(C1)), 3.25 (t, *J* = 7.3 Hz, 2 H, H-(C3), 2.33-2.21 (m, 2 H, H-(C2)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 121.6 (CN); 58.7 (C1); 37.4 (C3); 32.4 (C2).

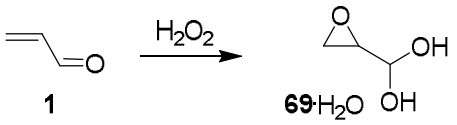
*Labelled **74***

The reaction described above (*Formation of 3-aminopropanal (**74**) from of acrolein (**1**) and ammonia*) was repeated using ^{15}N labelled ammonium chloride (38.1 mg, 0.7 mmol) at pH 6.5 and the reaction monitored by ^{1}H NMR over time (Appendix - *Figure 11.6*, *b*.).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 6.5)

3-Aminopropanal hydrate (^{15}N-**74**·H_2O): 5.17 (t, *J* = 5.6 Hz, 1 H, H-(C1)), 3.69 (t, *J* = 7.0 Hz, 2 H, H-(C3)), 1.96 (tdd, *J* = 7.0, 5.6, 2.9 Hz, 2 H, H-(C2)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 89.7 (C1); 58.8 (C3); 40.6 (C2).

*Formation of glycidaldehyde (**69**) from acrolein (**1**) with hydrogen peroxide*[287]

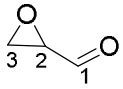


Acrolein (**1**; 9.3 µL, 0.14 mmol) was added to a stirred solution of hydrogen peroxide (13.2 µL, 10.6M, 0.14 mmol) in H_2O/D_2O (9:1, 1 mL) at RT and pH 8. The solution was adjusted to pH 8.5 with 1M NaOH and ^{1}H NMR spectra were periodically acquired, showing fast epoxidation to glycidaldehyde (**69**; 90% yield by ^{1}H NMR integration in <2 h).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 8.5) (*partial assignment*)

3-Hydroxypropanoic acid (**92**)[349]: 3.66 (t, *J* = 6.7 Hz, 2 H, H-(C2)); 2.31 (t, *J* = 6.7 Hz, 2 H, H-(C3)).

*Preparative synthesis of glycidaldehyde (**69**)*



Glycidaldehyde (**69**) was prepared by adaptation of Payne's synthesis[287] and Golding's continuous batch extraction[339]. To a solution of hydrogen peroxide (13.2 mL, 10.64$M_{(aq)}$, 0.14 mol) in water (95 mL) adjusted to pH 8, acrolein (**1**; 9.3 mL, 0.14 mol) was added dropwise over 10 min. The reaction was maintained between pH 8 and 8.5 (with 1M HCl/NaOH) and the solution was stirred for 2 h. The solution was then continuously extracted with CH_2Cl_2 for 48 h. The CH_2Cl_2 fraction was subsequently dried with $MgSO_4$ and filtered, giving a 34% yield (by ^{1}H NMR analysis) of extracted glycidaldehyde (**69**). The CH_2Cl_2 fraction containing the product **69** was distilled at 120 Torr. **69** was collected as a colourless liquid (2.74 g, 27%) and stored at -20°C under an argon atmosphere.

B.p. 62-64°C, 120 Torr; lit[287]. 57-58°C, 100 Torr. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.93 (d, *J* = 6.5 Hz, 1 H, H-(C1)); 3.34 (ddd, *J* = 6.5, 4.5, 2.5 Hz, 1 H, H-(C2)); 3.13 (dd, *J* = 5.3, 4.5 Hz, 1 H, H^a-(C3)); 3.03 (dd, *J* = 5.3, 2.5 Hz, 1 H, H^b-(C3)). ^{1}H NMR (600 MHz, D_2O[xxxvii]) δ 4.83 (d, *J* = 4.2 Hz, 1 H, H-(C1)); 3.11 (tdd, *J* = 4.2, 3.0, 0.7 Hz, 1 H, H-(C2)); 2.84 (td, *J* = 4.2, 0.7 Hz, 1 H, H^a-(C3)); 2.77 (dd, *J* = 4.2, 3.0 Hz, 1 H, H^b-(C3)). ^{13}C NMR (150 MHz, D_2O) δ_C 89.7 (C1); 54.4 (C2); 45.3 (C3). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_3H_5O_2^+$, 73.02895; found, 73.0288.

General procedure for the hydrolysis of glycidaldehyde (69**)**

Glycidaldehyde (**69**; 21 mg, 0.30 mmol) was dissolved in H_2O/D_2O (9:1, 0.75 mL; containing internal standard potassium hydrogen phtlatate (KHP), 26.2mM). The solution was adjusted to the specified pH (pH 3-9). The solution pH was then continuous monitored, adjusting as required with 1M NaOH/HCl, whilst ^{1}H NMR spectra were periodically acquired. Hydrolysis product glyceraldehyde (**12**) was confirmed by spiking with a commercial sample (Appendix - *Figure 11.7*).

Tables showing the results of the hydrolysis of glycidaldehyde (**69**). Comparison of pH on hydrolysis rate and formation of glyceraldehyde (**12**), represented in the *Figures 2.6*. The percentage of each compound in solution is based on ^{1}H NMR integration compared to the internal standard.

Time (h)	^{1}H NMR integration relative to starting concentration (%)	
	69	**12**
0.3	98	0
1	96	1
7	96	1
29	91	5
53	88	7
120	81	14
192	72	18

*Table 9.1: Hydrolysis reaction over time at pH 3 (**69** concentration: 400mM).*

[xxxvii] The data given in D_2O is for the hydrate of **69**, **69**·H_2O. Aldehydes glycidaldehyde (**69**), glyceraldehyde (**12**) and glycolaldehyde (**3**) sit as their hydrates in water. To simply we will refer to them without the additional "hydrate" addendum (·H_2O).

Time (h)	^{1}H NMR integration relative to starting concentration (%)	
	69	**12**
0.3	98	0
1	97	0
7	98	0
29	93	2
53	91	4
120	83	12
192	68	22

*Table 9.2: Hydrolysis reaction over time at pH 5 (**69** concentration: 400mM).*

Time (h)	^{1}H NMR integration relative to starting concentration (%)	
	69	**12**
0.3	98	0
1	97	0
7	96	0
29	92	5
53	90	7
120	65	16
192	60	26

*Table 9.3: Hydrolysis reaction over time at pH 7 (**69** concentration: 400mM).*

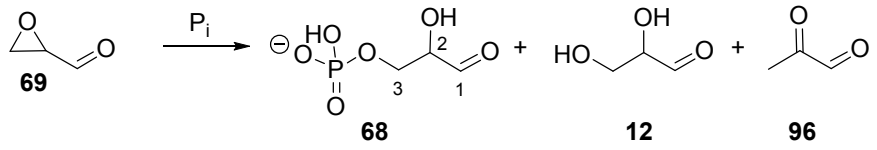
Time (h)	^{1}H NMR integration relative to starting concentration (%)	
	69	**12**
0.3	100	0
1	95	4
7	89	8
29	62	20
53	36	22
120	12	34

*Table 9.4: Hydrolysis reaction over time at pH 9 (**69** concentration: 400mM).*

9.3. Experimental Procedure for 3. Prebiotic incorporation of phosphate

General procedure for the phosphorylation of glycidaldehyde (69**)**

Glycidaldehyde (**69**; 18 mg, 0.25 mmol) was dissolved in P_i buffer (1 mL, 0.25-0.5M, pH 5-8) and incubated at RT. ^{1}H and ^{31}P NMR spectra were periodically acquired and production of glyceraldehyde-3-phosphate (**68**) was monitored. Glyceraldehyde (**12**; Appendix - *Figure 11.8*), and **68** synthesis were confirmed by sample spiking and corresponding signal amplification with a commercial standard. Methylglyoxal (**96**) was confirmed based on comparison with literature data[350].



69 **68** **12** **96**

Glyceraldehyde-3-phosphate (**68**): ^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 4.86 (d, J = 5.8 Hz, 1 H, H-(C1)); 3.77 (ddd, J = 11.3, 6.5, 3.3 Hz, 1 H, H^a-(C3)); 3.68 (dt, J = 11.3, 6.5 Hz, 1H, H^b-(C3)); 3.52 (td, J = 5.8, 3.3 Hz, 1 H, H-(C2)). ^{31}P (121 MHz, H_2O/D_2O 9:1) δ 6.41 (t, J = 6.5 Hz).

Tables showing the results of the phosphorylation of glycidaldehyde (**69**). Comparison of pH and quantity of P_i on formation of glyceraldehyde-3-phosphate (**68**) and total phosphorylation pathway product formation (**68** and its elimination to methylglyoxal (**96**)), represented in the *Figures 3.3, 3.5*. All percentages reported are proportional to the species integrated as shown in *Figure 3.4*. These add up to 85% when a standard is used.

Time (h)	^{1}H NMR integration relative to other discernible species integrated (%)				
	68	Total products: **68**+**96**	**69**	**96**	**12**
0.3	0	0	98.5	0	0
1.3	0	0	98.5	0	0
7.3	0.2	0.5	95.2	0.3	0.8
96	3.5	4.5	82.7	1	9.1
168	6.8	9.8	66.9	3	13.9

*Table 9.5: Phosphorylation reaction done at pH 5 with a 1:1 ratio of P_i:**69** (**69** concentration: 250mM).*

Time (h)	^{1}H NMR integration relative to other discernible species integrated (%)				
	68	Total products: **68**+**96**	**69**	**96**	**12**
0.3	0	0	98.5	0	0
1.3	0	0	98.5	0	0
7.3	2.4	2.6	93.1	0.2	0.8
96	12.1	31.9	56.1	19.9	6.1
168	14.9	41	28.3	26.1	8.7

*Table 9.6: Phosphorylation reaction done at pH 7 with a 1:1 ratio of P_i:**69** (**69** concentration: 250mM).*

Time (h)	^{1}H NMR integration relative to other discernible species integrated (%)				
	68	Total products: **68**+**96**	**69**	**96**	**12**
0.3	0.5	0.6	97.9	0.1	0.2
1.5	1.4	1.5	94.6	0.1	0.4
2.5	2.1	2.2	93.4	0.1	0.5
3.3	2.7	2.9	92.9	0.2	0.6
4.5	3.5	3.8	91.5	0.3	0.8
15.2	8.9	11.2	82.3	2.3	2.2
25.2	11.7	17.2	74.7	5.5	2.8
35.2	13	27.6	63.7	14.6	3.3
45	15	26.9	62.6	12	3.9
55.2	15.9	31.5	56.6	15.6	4.4
74	15.9	38.2	48.3	22.3	5.1

*Table 9.7: Phosphorylation reaction done at pH 7 with a 2:1 ratio of P_i:**69** (**69** concentration: 250mM).*

Time (h)	^{1}H NMR integration relative to other discernible species integrated (%)				
	68	Total products: **68**+**96**	**69**	**96**	**12**
0.3	0.8	0.8	95.6	0	0.2
1.5	2.2	2.3	93.3	0.1	0.5
2.5	3.3	3.5	92	0.2	0.6
3.3	4.3	4.7	90.7	0.4	0.7
4.5	5.3	5.9	89	0.6	1.1
5	5.5	8.4	86.9	2.9	1
15.2	13	16.2	76.6	3.3	2.4
25.2	15.9	24.5	66.5	8.6	3.5
35.2	18.7	33.2	56.8	14.4	4.1
45	19	44	45.2	25	4.8
55.2	21.3	45	43.1	23.7	5.3
74	21.3	53.3	32.8	32	5.9

*Table 9.8: Phosphorylation reaction done at pH 8 with a 2:1 ratio of P_i:**69** (**69** concentration: 250mM).*

Time (h)	^{1}H NMR integration relative to other discernible species integrated (%)				
	68	Total products: **68**+**96**	**69**	**96**	**12**
0.5	0.8	0.8	94.6	0	0.2
2	2.7	2.8	92.6	0.1	0.5
3.5	4.4	4.8	90.1	0.4	0.8
5	6.6	7.3	86.5	0.7	1.3
6.5	7.5	8.7	85.7	1.2	1.4
8	8.7	10.6	82.7	1.9	1.6
9.5	10	12.4	80.4	2.5	1.9
11	11.2	14.4	78	3.2	2.2
12.5	12.2	16.2	75.9	4	2.4
14	12.8	17.5	73.9	4.8	2.5
15.5	13.4	19.1	71.5	5.7	2.7
19.3	14.6	22.9	66.6	8.3	3.1
26	16.1	28.5	59.5	12.4	3.3
33.5	17.1	34.7	51	17.6	3.7
36	17.1	36.8	48.9	19.7	3.8
41	16.9	39.9	45	23	3.7
46	16.8	42.3	41.3	25.5	3.8
57.2	16.5	49.1	33.9	32.6	3.5
57.5	15.8	49.8	32.9	34	3.8
61	16	51.5	31.4	35.5	4

*Table 9.9: Phosphorylation reaction done at pH 7 with a 4:1 ratio of P_i:**69** (**69** concentration: 250mM).*

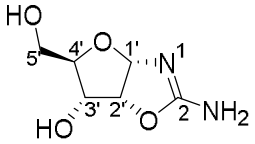
Time (h)	^{1}H NMR integration relative to other discernible species integrated (%)				
	68	Total products: **68**+**96**	**69**	**96**	**12**
0.5	1.1	1.3	92.1	0.2	0
2	3.1	3.3	90.8	0.2	0
3.5	5.5	5.9	88.2	0.5	0.9
5	8.3	9.5	79.6	1.2	1.8
6.5	16.9	32.9	50.3	16	3.1
8	16.7	39.5	47.9	22.8	2.3
9.5	16.3	49.7	29.5	33.4	3.4
11	14.1	53.7	22.1	39.6	3.6
12.5	11.8	58.4	16.3	46.6	3.3
14	11.5	60.7	14.9	49.2	2.5

*Table 9.10: Phosphorylation reaction done at pH 8 with a 4:1 ratio of P_i:**69** (**69** concentration: 250mM).*

9.4. Experimental Procedure for 4. Prebiotic synthesis of aminooxazolines through glycidaldehyde

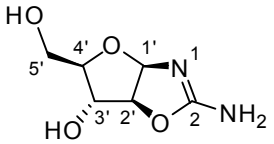
9.4.1. Syntheses of aminooxazolines (49) by action of cyanamide (4)

*D-Ribofuranosyl aminooxazoline (ribo-**49**)*[270]



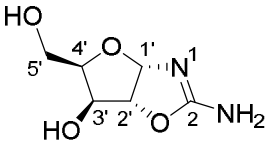
To a solution of D-ribose (**9**; 5.0 g, 33 mmol) in aqueous NH_4OH (3.5%, 5.3 mL) was added cyanamide (**4**; 4.0 g, 0.10 mol). The resulting solution was stirred at 60°C for 1 h. The reaction was cooled to RT, after which point methanol was added (11.1 mL). The mixture was cooled further to 4ºC and after 16 h white crystals were collected by filtration. The crystals were washed with ice-cold methanol and dried under vacuum to afford D-ribofuranosyl aminooxazoline (*ribo*-**49**) as a fine white powder (4 g, 80%). ^{1}H NMR (600 MHz, D_2O) δ 5.69 (d, J = 5.3 Hz, 1 H, H-(C1′)); 4.87 (t, J = 5.3 Hz, 1 H, H-(C2′)); 4.01 (dd, J = 9.5, 5.3 Hz, 1 H, H-(C3′)); 3.82 (ABX, J = 12.7, 2.4 Hz, 1 H, H^a-(C5′)); 3.62 (ABX, J = 12.7, 4.8 Hz, 1 H, H^b-(C5′)); 3.51 (ddd, J = 9.5, 4.8, 2.4 Hz, 1 H, H-(C4′)). ^{13}C NMR (151 MHz, D_2O) δ_C 166.7 (C2); 97.7 (C1′); 82.6 (C4′); 77.9 (C2′); 71.0 (C3′); 60.2 (C5′). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{11}N_2O_4^+$, 175.0713; found, 175.0604.

*D-Arabinofuranosyl aminooxazoline (arabino-**49**)*[24,xxxviii]



To a solution of D-arabinose (**13**; 1 g, 6.6 mmol) in methanol (3.3 mL) was added cyanamide (**4**; 0.56 g, 13.4 mmol) and a concentrated NH_4OH solution (0.32 mL). The resulting solution was stirred at 45°C for 5 h. Methanol was removed *in vacuo* and the residue was dissolved in water (2 mL) by heating. Cooling to RT yielded crystals which were collected by filtration. The crystals were washed with ice-cold H_2O (3 mL) and dried under vacuum to yield D-arabinofuranosyl aminooxazoline (*arabino*-**49**) as a fine white powder (640 mg, 55%). ^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 5.81 (d, J = 5.5 Hz, 1 H, H-(C1′)); 4.84 (d, J = 5.5 Hz, 1 H, H-(C2′)); 4.23 (d, J = 2.3 Hz, 1 H, H-(C3′)); 3.91 (dt, J = 4.1, 2.3, 2.3 Hz, 1 H, H-(C4′)); 3.50 (m, 1 H, H^a-(C5′)); 3.43 (m, 1 H, H^b-(C5′)). ^{13}C NMR (151 MHz, D_2O) δ_C 165.6 (C2); 99.3 (C1′); 89.4 (C2′); 85.0 (C4′); 75.9 (C3′); 61.8 (C5′). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{11}N_2O_4^+$, 175.0713; found, 175.0719.

*D-Xylofuranosyl aminooxazoline (xylo-**49**)*[306]



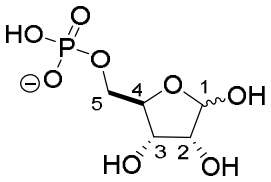
To a solution of D-xylose (**14**; 0.50 g, 3.3 mmol) in methanol (2 mL) was added cyanamide (**4**; 0.28 g, 6.7 mmol) and a concentrated NH_4OH solution (0.16 mL). The resulting solution was stirred at 60°C for 3 h. Methanol was removed *in vacuo* and the residue was dissolved in water (0.7 mL) by heating. Cooling to RT and further cooling to 4ºC for 36 h yielded crystals which were collected by filtration. The crystals were washed with ice-cold methanol (0.4 mL) and dried under vacuum to yield D-xylofuranosyl aminooxazoline (*xylo*-**49**) as a fine white powder (62 mg, 11%). ^{1}H NMR (600 MHz, D_2O) δ 5.84 (d, J = 5.3 Hz, 1 H, H-(C1′)); 4.79 (d, J = 5.3 Hz, 1 H, H-(C2′)); 4.27 (d, J = 2.6 Hz, 1 H, H-(C3′)); 3.80

[xxxviii] Prepared in the Powner lab by A. Nikmal

(ABX, J = 11.3, 4.9 Hz, 1 H, H^a-(C5′)); 3.77 (m, 1 H, H-(C4′)); 3.70 (ABX, J = 11.3, 6.4 Hz, 1 H, H^b-(C5′)). ^{13}C NMR (151 MHz, D_2O) δ_C 165.9 (C2); 98.2 (C1′); 87.8 (C2′); 79.2 (C4′); 74.0 (C3′); 59.7 (C5′). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{11}N_2O_4^+$, 175.0713; found, 175.0947.

9.4.2. Syntheses of aminooxazolines 5′-phosphate (53) by action of cyanamide (4)

*D-Ribose 5-phosphate (ribo-**52**)*[351,xxxix]

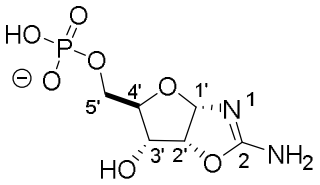


In a suspension of Dowex® 50W×8 resin (10 g, H^+-form, prewashed) in water (20 mL) was added β-D-adenosine-5′-phosphate (*ribo*-**152**; 1.50 g, 4.33 mmol). The solution was heated to 100°C and refluxed for 30 min. The reaction mixture was allowed to cool to RT and the resin removed by filtration. The resin was washed with water. The combined solution was then adjusted to pH 7.5 at 0°C with 4M NaOH soln. to obtain the disodium salt. The solution was lyophilised to afford compound D-ribose 5-phosphate (*ribo*-**52**; 0.88 g, 74 %) as a white solid.

β anomer

^{1}H NMR (600 MHz, D_2O) δ 5.31 (d, *J* = 2.0 Hz, 1 H, H-(C1)); 4.37 (dd, *J* = 5.2, 4.9 Hz, 1 H, H-(C3)); 4.16 (ddd, *J* = 5.2, 4.5, 4.4 Hz, 1 H, H-(C4)); 4.12 (dd, *J* = 4.9, 2.0 Hz, 1 H, H-(C2)); 4.06 (ABXY, *J* = 11.2, 4.5 Hz, 1 H, H^a-(C5)); 4.01-3.93 (ABXY, 1 H, H^b-(C5)). ^{13}C NMR (151 MHz, D_2O) δ 101.8 (C1), 82.4 (C4), 76.0 (C2), 71.3 (C3), 65.8 (C5). HRMS (ESI^-) (*m/z*): $[M-H]^-$ calcd for formula $C_5H_{10}O_8P^-$, 229.0119; found, 229.0112.

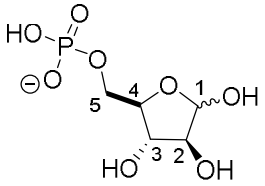
*D-Ribofuranosyl aminooxazoline 5′-phosphate (ribo-**53**)*[283,xxxix]



[xxxix] Prepared in the Powner lab by Dr. C. A. Fernández García.

Cyanamide (**4**; 1.25 g, 29.7 mmol) was added to a solution of D-ribose 5-phosphate (*ribo*-**52**; 3.58 g, 13.1 mmol) in aqueous NH_4OH (3.5%, 40 mL). The resulting solution was stirred at 60°C for 3 h and was then lyophilised. The reaction was purified *via* ion exchange chromatography with Dowex® 50W×8 resin (40 g, H^+-form, prewashed, eluting with water (100 mL), HCl 0.125M (100 mL), 0.25M (100 mL), 0.5M (100 mL), 1M (100 mL) and 2M (100 mL). Expected product D-ribofuranosyl aminooxazoline 5′-phosphate (*ribo*-**53**) was found in fractions 0.125M HCl and/or 0.25M HCl, confirmed by ^{1}H-NMR. Lyophilised fractions containing *ribo*-**53** were combined to yield *ribo*-**53** a yellowish solid (2.33 g, 70% isolated yield). ^{1}H NMR (600 MHz, D_2O) δ 5.97 (d, J = 5.2 Hz, 1 H, H-(C1′)); 5.37 (t, J = 5.2 Hz, 1 H, H-(C2′)); 4.28 (dd, J = 9.3, 5.2 Hz, 1 H, H-(C3′)); 4.14 (ABX, J = 12.0, 5.6, 2.0 Hz, 1 H, H^a-(C5′)); 3.98 (ABX, J = 12.0, 7.0, 4.5 Hz, 1 H, H^b-(C5′)); 3.91 (dt, J = 9.3, 2.0 Hz, 1 H, H-(C4′)). ^{13}C NMR (151 MHz, D_2O) δ 164.0 (C2); 88.5 (C1′); 86.0 (C2′); 78.3 (d, C4′); 70.0 (C3′); 63.7 (d, C5′). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled) δ 0.04.

*D-Arabinose 5-phosphate (arabino-**52**)*[351]



In a suspension of Dowex® 50W×8 resin (5 g, H^+-form, prewashed) in water (10 mL) was added adenine 9-β-D-arabinofuranoside-5′-phosphate (*arabino*-**152**; 0.52 g, 1.5 mmol). The solution was heated to 100°C and refluxed for 30 min. The reaction mixture was allowed to cool to RT and the resin removed by filtration. The resin was washed with water. The combined solution was then adjusted to pH 7.5 with 4M NaOH. The solution was lyophilised to afford compound D-arabinose 5-phosphate (*arabino*-**52**; 378 mg, quant. yield) as a white solid.

β anomer

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 5.19 (d, J = 2.8 Hz, 1 H, H-(C1)); 4.12 (dd, J = 5.4, 3.7 Hz, 1H, H-(C4)); 4.02-3.98 (1 H, H-(C3), *partially obscured*);

3.97-3.93 (1 H, H-(C2), *partially obscured*); 3.93-3.84 (2 H, H-(C5), *partially obscured*). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 101.8 (C1); 82.5 (C4); 82.0 (C2); 76.2 (C3); 66.0 (C5). ^{31}P NMR (162 MHz, H_2O/D_2O 9:1, 1H-decoupled): δ 1.8. HRMS (ESI^+) (*m/z*): $[M+Na^+]^+$ calcd for formula $C_5H_{11}NaO_8P^+$, 253.0084; found, 253.0087.

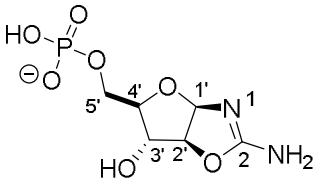
α anomer

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 5.21 (d, *J* = 4.5 Hz, 1 H, H-(C1)); 4.08-4.04 (1 H, H-(C3), *partially obscured*); 4.04-4.01 (1 H, H-(C2), *partially obscured*); 3.93-3.84 (2 H, H-(C5), *partially obscured*); 3.84-3.81 (1 H, H-(C4), *partially obscured*). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 95.9 (C1); 80.8 (C4); 76.8 (C2); 74.8 (C3); 64.8 (C5).

Open chain

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 4.97 (d, *J* = 7.0 Hz, 1 H, H-(C1)).

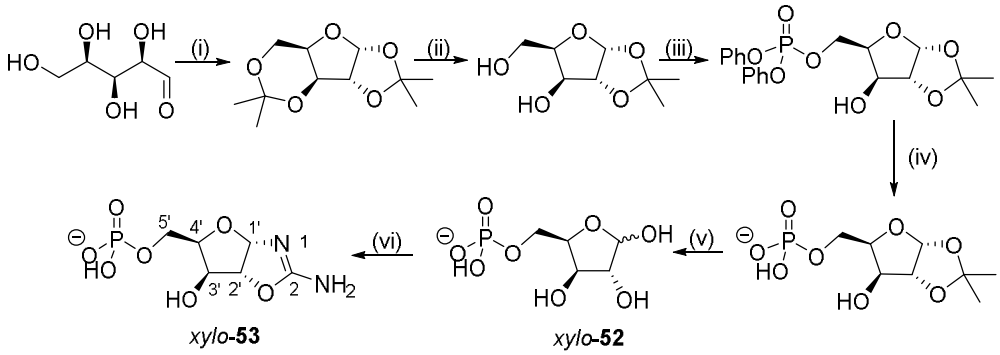
*D-Arabinofuranosyl aminooxazoline 5′-phosphate (arabino-**53**)*[283]



Cyanamide (**4**; 378 mg, 9 mmol) was added to a solution of D-arabinose 5-phosphate (*arabino*-**52**; 1.13g, 4.5 mmol) in water (5 mL). The resulting solution was stirred at 60°C for 4 h and was then lyophilised. The reaction was purified *via* ion exchange chromatography with Dowex® 50W×8 resin (40 g, H^+-form, prewashed, eluting with water (100 mL), HCl 0.125M (100 mL), 0.25M (100 mL), 0.5M (100 mL), 1M (100 mL) and 2M (100 mL). Expected product D-arabinofuranosyl aminooxazoline 5′-phosphate (*arabino*-**53**) was found in fractions 0.25M HCl and/or 0.5M HCl, confirmed by ^{1}H-NMR. Lyophilised fractions containing *arabino*-**53** were combined to yield *arabino*-**53** as a yellowish solid (827 mg, 72% isolated yield). ^{1}H NMR (600 MHz, D_2O) δ 6.11 (d, *J* = 5.7 Hz, 1 H, H-(C1′)); 5.37 (d, *J* = 5.7 Hz, 1 H, H-(C2′)); 4.61 (d, *J* = 2.3 Hz, 1 H, H-(C3′)); 4.31 (dt, *J* = 4.1, 2.3 Hz, 1 H, H-(C4′)); 3.86-3.82 (m, 2 H, H-(C5′)). ^{13}C

NMR (151 MHz, D_2O) δ_C 163.2 (C2); 92.8 (C2′); 90.3 (C1′); 86.8 (C4′); 75.0 (C3′); 65.4 (C5′). ^{31}P NMR (162 MHz, D_2O, 1H-decoupled): δ 0.2. HRMS (ESI^+) (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{12}N_2O_7P^+$, 255.0377; found, 255.0375.

*D-Xylofuranosyl aminooxazoline 5′-phosphate (xylo-**53**)*[352,xl]



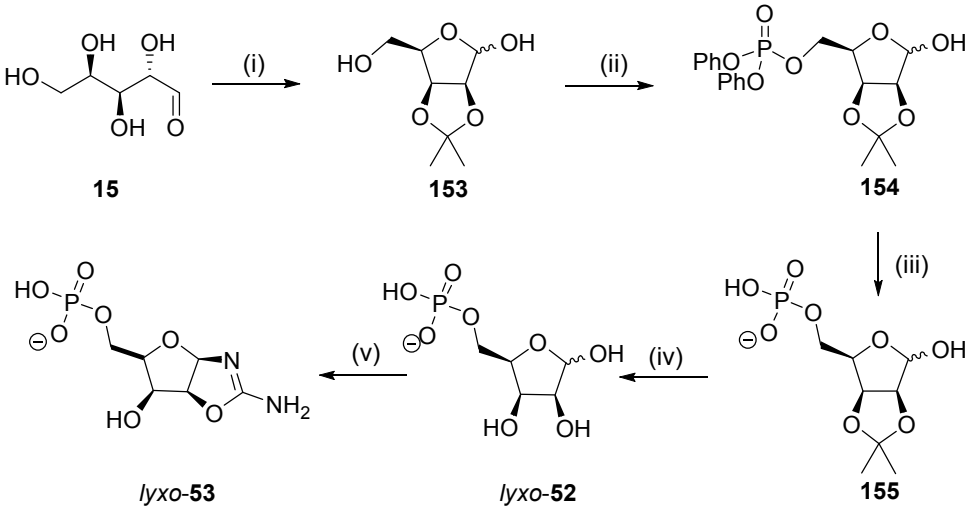
Scheme 9.1: Reagents and conditions: (i) $CuSO_4$, H_2SO_4, $(CH_3)_2CO$, 4 days, RT; (ii) 0.1M HCl_{aq}, 3 h, RT, 44% over two steps; (iii) $(PhO)_2POCl$, CH_2Cl_2, Py, 7 h, 0°C, 90%; (iv) H_2, PtO_2, MeOH, 16 h, RT, 97%; (v) H_2O, 16 h, 50°C, 87%; (vi) H_2NCN, NH_4OH, H_2O, 64%.

Cyanamide (**4**; 212 mg, 5.04 mmol) was added to a solution of D-xylose 5-phosphate (*xylo*-**52**; 580 mg, 2.3 mmol) in aqueous NH_4OH (3.5%, 8 mL). The resulting solution was stirred at 60°C for 4 h and was then lyophilised. The reaction was purified *via* ion exchange chromatography with Dowex® 50W×8 resin (40 g, H^+-form, prewashed, eluting with water (100 mL), HCl 0.125M (100 mL), 0.25M (100 mL), 0.5M (100 mL), 1M (100 mL), 2M (100 mL) and 2M (100 mL). Expected product D-xylofuranosyl aminooxazoline 5′-phosphate (*xylo*-**53**) was found in fractions 0.125M HCl and/or 0.25M HCl, confirmed by ^1H-NMR. Lyophilised fractions containing *xylo*-**53** were combined to yield *xylo*-**53** as a colourless solid (375 mg, 64% isolated yield). ^1H NMR (600 MHz, D_2O) δ 6.11 (d, *J* = 5.2 Hz, 1 H, H-(C1′)), 5.30 (d, *J* = 5.2 Hz, 1 H, H-(C2′)), 4.52 (d, *J* = 2.9 Hz, 1 H, H-(C3′)), 4.21 (ddd, *J* = 7.1, 4.5, 2.9 Hz, 1 H, H-(C4′)), 4.11 (ABX, *J* = 11.3, 7.1, 4.5 Hz, 1 H, H^a-(C5′)), 3.99 (ABX, *J* = 11.3, 7.1 Hz, 1 H, H^b-(C5′)). ^{13}C NMR (151 MHz, D_2O) δ 163.4 (C2), 90.8 (C2′), 89.1 (C1′), 79.8 (d, C4′), 72.8

[xl] Prepared in the Powner lab by Dr. C. A. Fernández García. Only the synthesis and data for the final compound *xylo*-**53** is given here, as it is used for spiking.

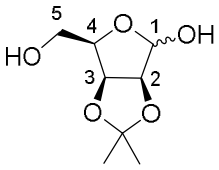
(C3′), 63.1 (d, C5′). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled) δ 0.61. HRMS (ESI) (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{12}N_2O_7P^+$, 255.0377; found, 255.0375.

Synthesis of D-lyxose aminooxazoline 5′-phosphate (*lyxo*-53)



Scheme 9.2: Reagents and conditions: (i) H_2SO_4, $(CH_3)_2CO$, 2 h, RT, quant. yield; (ii) $(PhO)_2POCl$, CH_2Cl_2, Py, 6 h, 0°C, 57%; (iii) H_2, PtO_2, MeOH, 18 h, RT, quant. yield; (iv) H_2O, 16 h, 50°C, quant. yield; (v) H_2NCN, NH_4OH, H_2O.

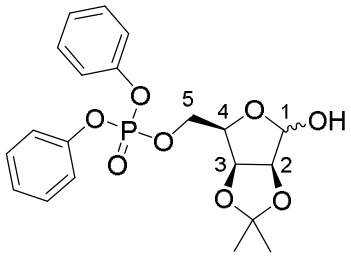
2,3-O-*Isopropylidene-D-lyxofuranose (**153**)*[353]



2,3-*O*-Isopropylidene-D-lyxofuranose (**153**) was prepared following a procedure adapted from Tsai *et al.*[353] D-lyxose (**15**; 584 mg, 3.9 mmol) was suspended in dry acetone (19.5 mL) and the resulting solution was cooled to 0°C. Concentrated sulfuric acid (22 μL, 0.39 mmol) was added dropwise. The solution was stirred at 0°C for 10 min then warmed to RT and stirred for a further 2 h. Aqueous NH_4OH solution (3.5 %) was subsequently added to neutralise the solution. The solution was then filtered and concentrated *in vacuo* to yield 2,3-*O*-isopropylidene-D-lyxofuranose (**153**) as an oil (812 mg, quant. yield).

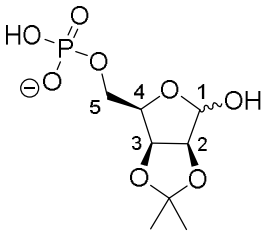
$[\alpha]_D^{25}$= +21.25 (c=0.16, Acetone). IR (cm^{-1}) 3630-3120 (O-H), 3060-2810 (C-H).^{1}H NMR (600 MHz, $CDCl_3$) δ 5.44 (s,1 H, H-(C1)); 4.82 (dd, *J* = 5.8, 3.9 Hz, 1 H, H-(C3)); 4.63 (d, *J* = 5.8 Hz, 1 H, H-(C2)); 4.30-4.26 (m, 1 H, H-(C4)); 3.96-3.92 (m, 2 H, H(C5)); 1.47 (s, 1H, $C(CH_3)_2$); 1.32 (s, 1H, $C(CH_3)_2$). ^{13}C NMR (150 MHz, $CDCl_3$) δ_C 112.7 ($\underline{C}(CH_3)_2$); 100.9 (C1); 85.6 (C2); 80.4 (C3); 79.7 (C4); 61.1 (C5), 26.0 ($C(\underline{C}H_3)_2$), 24.5 ($C(\underline{C}H_3)_2$). HRMS (ES^+) (*m/z*): $[M+NH_4^+]^+$ calcd for formula $C_8H_{18}O_5N^+$, 208.1179; found, 208.1180.

2,3-O-*Isopropylidene-D-5-diphenylphospholyxofuranose* ***(154)***



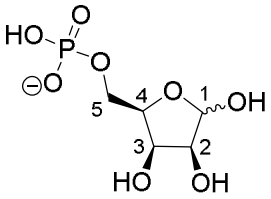
2,3-*O*-Isopropylidene-D-lyxofuranose (**153**; 1.085 g, 5.7 mmol) was dissolved in a mixture of anhydrous CH_2Cl_2 (270 mL) and anhydrous pyridine (40 mL). The resulting solution was cooled to 0°C and diphenyl chlorophosphate (1.49 mL, 7.15 mmol) added dropwise and stirred at 0°C. After 6 h the solution was warmed to RT, and quenched by addition of water (40 mL) and the solution was concentrated *in vacuo*. The residue was purified by column chromatography (EtOAc/Hexane 20% to 100% EtOAc) to yield 2,3-*O*-isopropylidene-D-5-diphenylphospholyxofuranose (**154**) as an oil (1.36 g, 57%). $[\alpha]_D^{25}$= +16.90 (c=0.073, $CHCl_3$). IR (cm^{-1}) 3560-3250 (O-H), 3120-2810 (C-H), 1175 (P=O). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.36-7.31 (m, 4 H, Ph); 7.25-7.22 (m, 4 H, Ph); 7.21-7.17 (m, 2 H, Ph); 5.39 (s, 1 H, H-(C1)); 4.74 (dd, *J* = 5.8, 3.4 Hz, 1 H, H-(C3)); 4.60 (d, *J* = 5.8 Hz, 1 H, H-(C2)); 4.57-4.53 (m, 1 H, H^a-(C5)); 4.45-4.40 (m, 2H, H-(C4) and H^b-(C5)); 1.43 (s, 1H, $C(CH_3)_2$); 1.28 (s, 1H, $C(CH_3)_2$). ^{13}C NMR (150 MHz, $CDCl_3$) δ_C 150.4 (Ph); 129.7 (Ph); 125.4 (Ph); 120.1 (Ph); 112.8 ($\underline{C}(CH_3)_2$); 101.2 (C1); 85.4 (C2); 79.5 (C3); 78.3 (C4); 67.1 (C5), 25.9 ($C(\underline{C}H_3)_2$), 24.6 ($C(\underline{C}H_3)_2$). ^{31}P NMR (121 MHz, $CDCl_3$, 1H-decoupled): δ -12.1. HRMS (ESI) (*m/z*): $[M+H^+]^+$ calcd for formula $C_{20}H_{24}O_8P^+$, 423.1203; found, 423.1204.

2,3-O-*Isopropylidene-D-5-phospholyxofuranose (**155**)*



2,3-*O*-Isopropylidene-D-5-diphenylphospholyxofuranose (**154**; 280 mg, 0.66 mmol) was dissolved in methanol (14 mL) under an N_2 atmosphere, and PtO_2 (25 mg, catalytic) was added. The suspension was stirred vigorously overnight under an H_2 atmosphere for 16 h. The reaction was then flushed with N_2 and the solution was decanted off the catalyst and filtered through a Celite® pad, followed by washing with methanol (3×5 mL). The combined filtrates were concentrated *in vacuo* to yield 2,3-*O*-isopropylidene-D-5-phospholyxofuranose (**155**) as a colourless oil (170 mg, quant. yield). $[\alpha]_D^{25}$= +1.55 (c=0.344, CH_3OH). IR (cm^{-1}) 3500-3130 (O-H), 3000-2810 (C-H), 1195 (P=O). ^{1}H NMR (600 MHz, CD_3OD) δ 5.24 (s, 1 H, H-(C1)); 4.81 (dd, *J* = 5.8, 3.8 Hz, 1 H, H-(C3)); 4.55 (d, *J* = 5.8 Hz, 1 H, H-(C2)); 4.35-4.31 (m, 1 H, H-(C4)); 4.23 (ABXY, *J* = 10.7, 6.8, 4.8 Hz, 1 H, H^a-(C5)); 4.08 (ABXY, *J* = 10.7, 6.8 Hz, 1 H, H^b-(C5)); 1.40 (s, 1H, $C(CH_3)_2$); 1.29 (s, 1H, $C(CH_3)_2$). ^{13}C NMR (150 MHz, CD_3OD) δ_C 113.8 ($\underline{C}(CH_3)_2$); 102.4 (C1); 87.3 (C3); 81.2 (C2); 80.0 (C4); 65.9 (C5), 26.5 ($C(\underline{C}H_3)_2$), 25.0 ($C(\underline{C}H_3)_2$).^{31}P NMR (162 MHz, CD_3OD, 1H-decoupled): δ 0.1. HRMS (ES^-) (*m/z*): $[M]^-$ calcd for formula $C_8H_{14}O_8P^-$, 269.0432; found, 269.0435.

*D-Lyxose 5-phosphate (lyxo-**52**)*[283,351]



2,3-*O*-Isopropylidene-D-5-phospholyxofuranose (**155**; 170 mg, 0.66 mmol) was dissolved in water (25 mL) and heated at 50°C for 16 h. The solution was then adjusted to pH 6.5 and then concentrated *in vacuo* to yield D-lyxose 5-phosphate (*lyxo*-**52**) as a colourless oil (145 mg, quant.).

β anomer

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 5.22 (d, *J* = 4.8 Hz, 1 H, H-(C1)); 4.37-4.31 (m, 1 H, H-(C4)); 4.26 (t, *J* = 3.9 Hz, 1 H, H-(C3)); 4.05-4.01 (m, 1 H, H-(C2), *peak overlaps with α anomer H^{a}-(C5)*); 3.96 (ABXY *J* = 10.9, 6.7 Hz, 1H, H^{a}-(C5)); 3.83 (ABXY, *J* = 10.9, 6.7 Hz, 1 H, H^{b}-(C5)). ^{13}C NMR (150 MHz, H_2O+D_2O) δ_C 101.4 (C1); 79.8 (C4); 77.7 (C2); 71.7 (C3); 63.9 (C5). ^{31}P NMR (162 MHz, H_2O/D_2O 9:1, 1H-decoupled): δ 2.1. HRMS (ES^-) (*m/z*): $[M]^-$ calcd for formula $C_5H_{10}O_8P^-$, 229.0119; found, 229.0118.

α anomer

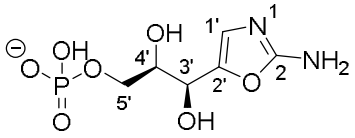
^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 5.18 (d, *J* = 4.9 Hz, 1 H, H-(C1)); 4.23 (t, *J* = 4.1 Hz, 1 H, H-(C3)); 4.13-4.10 (m, 1 H, H-(C2)); 4.10-4.08 (m, 1 H, H-(C4)); 4.05-4.01 (m, 1 H, H^{a}-(C5), *peak overlaps with β anomer H-(C2)*); 3.89 (ABXY, *J* = 10.9, 6.9 Hz, 1 H, H^{b}-(C5)). ^{13}C NMR (150 MHz, H_2O/D_2O 9:1) δ_C 96.2 (C1); 80.0 (C4); 72.0 (C2); 70.4 (C3); 64.6 (C5). ^{31}P NMR (162 MHz, H_2O/D_2O 9:1, 1H-decoupled): δ 2.1.

Open chain isomer (*partial assignment*)

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ 4.91 (d, *J* = 4.1 Hz, 1 H, H-(C1)); 4.08-4.04 (m, 1 H, H-(C2)). ^{13}C NMR (150 MHz, H_2O/D_2O 9:1) δ_C 76.7 (C2), 71.6 (C1).

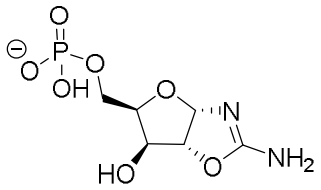
Reaction of D-lyxose 5-phosphate (*lyxo*-52) with cyanamide (4)

*Forming aromatic intermediate (2R,3S)-3-(2-aminooxazole-5-yl)-2,3 dihydroxypropyl phosphate **100**a*



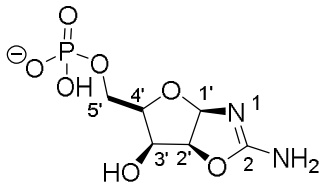
D-Lyxose 5-phosphate (*lyxo*-**52**; 116 mg, 0.5 mmol) and cyanamide (**4**; 42 mg, 1 mmol) were dissolved in H_2O/D_2O (9:1, 0.5 mL) and the solution was adjusted to pH 7. The resulting solution was heated at 60°C for 3 h (*Figure 6.11*).

*Forming xylose aminooxazoline 5′-phosphate (xylo-**53**)*



D-Lyxose 5-phosphate (*lyxo*-**52**; 6 mg, 0.022 mmol) and cyanamide (**4**; 5 mg, 0.12 mmol) were dissolved in H_2O/D_2O (9:1, 1 mL) at the solution was adjusted to pH 7. The resulting solution was heated at 60°C for 3 h, and subsequently spiked with a synthesised sample of *xylo*-**53** (*Figure 6.12*).

*Forming lyxose aminooxazoline 5′-phosphate (lyxo-**53**)*



D-Lyxose 5-phosphate (*lyxo*-**52**; 11.6 mg, 0.05 mmol) and cyanamide (**4**; 4.2 mg, 0.1 mmol) were dissolved in an aqueous NH_4OH solution (3%, 0.5 mL, H_2O/D_2O 9:1). The resulting solution was incubated at RT and at 60°C. Reactions were analysed by ^{1}H NMR periodically (*Figures 6.13 and 6.14*).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) for the RT reaction.

2-Aminooxazole (**55**): 7.36 (s, 1 H, H-(C4)); 6.83 (s, 1 H, H-(C5)).

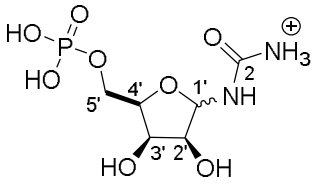
(2*R*,3*S*)-3-(2-Aminooxazole-5-yl)-2,3-dihydroxypropyl phosphate (**100**a): 6.76 (s, 1 H, H-(C1′)).

D-Lyxofuranosyl aminooxazoline-5′-phosphate (*lyxo*-**53**): 5.77 (d, J = 5.8 Hz, 1 H, H-(C1′)); 5.11 (t, J = 5.8 Hz, 1 H, H-(C2′)); 4.59 (t, J = 5.8 Hz, 1 H, H-(C3′)); 4.19 (dt, J = 7.2, 5.8 Hz, 1 H, H-(C4′)); 4.04-3.99 (m, 1 H, H^a-(C5′)); 3.82-3.70 (m, 1 H, H^b-(C5′)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 166.2 (C2); 98.9 (C1′); 83.3 (C2′); 80.0 (C4′); 71.4 (C3′); 63.1 (C5′). ^{31}P NMR (121 MHz, H_2O/D_2O 9:1,

1H-decoupled): δ 4.57. HRMS (ESI^+) (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{12}N_2O_7P^+$, 255.0377; found, 255.0377.

By-product **108**: 4.88 (d, *J* = 3.9 Hz, 1 H, H-(C1′)); 4.34 (dd, *J* = 7.2, 3.9 Hz, 1 H, H-(C2′)); 3.99 (m, 1 H, H-(C4′)); 3.91 (m, 2 H, H-(C5′)); 3.83 (m, 1 H, H-(C3′)). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 163.8 (C2); 86.6 (C2′); 74.8 (C1′); 70.7 (C3′); 70.7 (C4′); 65.6 (C5′). ^{31}P NMR (121 MHz, H_2O/D_2O 9:1, 1H-decoupled): δ 4.61. HRMS (ESI^+) (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{12}N_2O_7P^+$, 273.0482; found, 273.0484.

Proposed structure of **108**:



Formation of pentose aminooxazolines (49**) from glyceraldehyde (**12**) and 2-aminooxazole (**55**)**

Pentose aminooxazolines (**49**) were prepared following a procedure adapted from Anastasi *et al.*[272]. 2-aminooxazole (**55**; 14 mg, 0.16 mmol) was dissolved in H_2O/D_2O (9:1, 1 mL; containing internal standard KHP, 24.2mM) and glyceraldehyde (**12**; 12.52 mg, 0.14 mmol) was added. The solution was adjusted to pH 7 and the reaction monitored by ^1H NMR periodically. After 6 days at RT, the reaction showed a 73% yield of pentose aminooxazolines (**49**), which was confirmed by a positive spike with an authentic standard of *xylo*-**49**.

^1H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

2-aminooxazole (**55**): 7.20 (s, 1 H, H-(C4)); 6.67 (s, 1 H, H-(C5)).

(2*S*)-1-(4-hydroxy-2-iminooxazolidin-5-yl)propane-1,2,3-triol (**99**)[xli]: 5.34 (d, *J* = 5.5 Hz, 1 H, H-(C1′)).

rac-Ribofuranosyl aminooxazoline (*ribo*-**49**): 5.70 (d, *J* = 5.0 Hz, 1 H, H-(C1′)).

rac-Arabinofuranosyl aminooxazoline (*arabino*-**49**): 5.81 (d, *J* = 5.5 Hz, 1 H, H-(C1′)).

rac-Xylofuranosyl aminooxazoline (*xylo*-**49**): 5.84 (d, *J* = 5.1 Hz, 1 H, H-(C1′)).

rac-Lyxofuranosyl aminooxazoline (*lyxo*-**49**f): 5.64 (d, *J* = 5.6 Hz, 1 H, H-(C1′)).

rac-Lyxopuranosyl aminooxazoline (*lyxo*-**49**p): 5.41 (d, *J* = 5.8 Hz, 1 H, H-(C1′)).

Formation of 5′-phosphorylated pentose aminooxazolines (53**) from glyceraldehyde-3-phosphate (**68**) and 2-aminooxazole (**55**)**

Aminooxazolines 5′-phosphate (**53**) were prepared following a procedure adapted from Anastasi *et al.*[283]. 2-aminooxazole (**55**; 5.9 mg, 0.07 mmol) was dissolved in H_2O/D_2O (9:1, 0.77 mL; containing KHP, 14.13mM) and glyceraldehyde-3-phosphate (**68**; 235 μL of 50 mg/mL aqueous solution, 0.07 mmol) was added. The solution was adjusted to pH 7 and monitored, re-adjusted periodically (pH fluctuations ± 0.2 were observed). The reaction was monitored by ^{1}H NMR periodically to show a 23% yield of aminooxazolines 5′-phosphate (**53**) after 93 h at RT. Presence of **53** was confirmed by ^{31}P-^{1}H HMBC.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

2-aminooxazole (**55**): 7.21 (d, *J* = 1 Hz, 1 H, H-(C4)); 6.68 (d, *J* = 1 Hz, 1 H, H-(C5)).

(2*R*)-3-(2-Aminooxazole-5-yl)-2,3-dihydroxypropyl phosphate (**100**)[xlii]: 6.66 (s, 1 H, H-(C1′)).

[xli] Only one of the 8 isomers of **99** was detectable.
[xlii] There is one detectable isomer out of the possible 2 isomers of **100**.

(2*S*)-2,3-dihydroxy-3-(4-hydroxy-2-iminooxazolidin-5-yl)propyl phosphate (**101**)[xliii]:

5.68 (d, J = 2.1 Hz, 1 H, H-(C1′)).

5.66 (d, J = 2.1 Hz, 1 H, H-(C1′)).

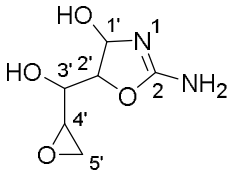
rac-Ribofuranosyl aminooxazoline 5′-phosphate (*ribo*-**53**): 5.84 (d, J = 5.1 Hz, 1 H, H-(C1′)).

rac-Arabinofuranosyl aminooxazoline 5′-phosphate (*arabino*-**53**): 5.93 (d, J = 5.5 Hz, 1 H, H-(C1′)).

rac-Xylofuranosyl aminooxazoline 5′-phosphate (*xylo*-**53**): 5.98 (d, J = 5.3 Hz, 1 H, H-(C1′)).

9.4.3. Glycidaldehyde (69) and 2-aminooxazole (55) experiments

Preparative synthesis of oxazolines (97)



2-Aminooxazole (**55**; 70.6 mg, 0.84 mmol) was dissolved in H_2O/D_2O (9:1, 5 mL, containing KHP, 24.2mM) and glycidaldehyde (**69**; 50.4 mg, 0.63 mmol[xliv]) was added. The solution was adjusted to pH 5 using 1M HCl, monitored and re-adjusted periodically (pH fluctuations ±0.8 were observed). ^{1}H NMR spectra were periodically acquired and after 13 h at RT a maximum yield of oxazoline **97** (68% with respect to initial **69** concentration, assessed by ^{1}H NMR, by integrating the H-C1′ proton resonances relative to internal standard).

[xliii] There are 2 detectable isomers out of the possible 8 isomers of **101**. Data is given for two that were assignable.

[xliv] The purity of **69** decreased over time to 90% (by ^{1}H NMR analysis with a standard), leading to the observed discrepancy between the weight added and the moles added.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 5) (*partial assignment*)

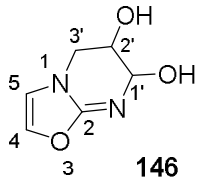
(1*S*)-((*S*)-2-Amino-4,5-dihydrooxazol-5-yl)((oxiran-2-yl)methanol (**97**a): 5.69 (d, *J* = 2.6 Hz, 1 H, H-(C1′)); 4.88 (dd, *J* = 4.4, 2.6 Hz, 1 H, H-(C2′)); 4.05 (dd, *J* = 4.4, 3.4 Hz, 1 H, H-(C3′)); 3.26 (dd, *J* = 4.2, 3.4, 3.1 Hz, 1 H, H-(C4′)); 2.95-2.75 (m, 2 H, H-(C5′)[xlv]). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 163.1 (C2); 91 (C2′); 80.5 (C1′); 68.0 (C3′); 52.3 (C4′).

(1*R*)-((*R*)-2-Amino-4,5-dihydrooxazol-5-yl)((oxiran-2-yl)metanol (**97**b): 5.67 (d, *J* = 2.8 Hz, 1 H, H-(C1′)); 4.83 (dd, *J* = 3.9, 2.8 Hz, 1 H, H-(C2′)); 4.02 (dd, *J* = 4.3, 3.9 Hz, 1 H, H-(C3′)); 3.20 (td, *J* = 4.3, 4.3, 2.9 Hz, 1 H, H-(C4′)); 2.95-2.75 (m, 2 H, H-(C5′)[xlv]). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 163.1 (C2); 90.8 (C2′); 80.6 (C1′); 68.8 (C3′); 51.6 (C4′).

(1*R*)-((*S*)-2-Amino-4,5-dihydrooxazol-5-yl)((oxiran-2-yl)metanol (**97**c): 5.63 (d, *J* = 2.7 Hz, 1 H, H-(C1′)); 4.91 (t, *J* = 2.7 Hz, 1 H, H-(C2′)); 3.86 (dd, *J* = 5.5, 2.7 Hz, 1 H, H-(C3′)); 3.22 (m, 1 H, H-(C4′)); 2.95-2.75 (m, 2 H, H-(C5′)[xlv]). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 163.4 (C2); 90.7 (C2′); 81.6 (C1′); 69.8 (C3′); 51.8 (C4′).

(1*S*)-((*R*)-2-Amino-4,5-dihydrooxazol-5-yl)((oxiran-2-yl)metanol (**97**d): 5.62 (d, *J* = 2.7 Hz, 1 H, H-(C1′)); 4.93 (t, *J* = 2.7 Hz, 1 H, H-(C2′)); 3.92 (dd, *J* = 4.6, 2.7 Hz, 1 H, H-(C3′)); 3.28-3.26 (m, 1 H, H-(C4′), *peak overlaps with H-(C4'*) *of other isomer*); 2.95-2.75 (m, 2 H, H-(C5′)[xlv]). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 163.4 (C2); 91.3 (C1′); 81.8 (C2′); 69.8 (C3′); 53.0 (C4′).

Minor peaks were detected and assigned to hemi-aminal **146** (*partial assignment*)



146

[xlv] All H-(C5') overlap at 2.95-2.75 ppm; ^{13}C was assigned by ^{13}C NMR and HSQC; however, it was not possible to assign peaks for C5' for every stereoisomer as all the peaks appear in the same spectral region 46.4-44.5 ppm.

Oxazole **146**: 7.57 (d, *J* = 1.7 Hz, 1 H, H-(C5)); 7.34 (d, *J* = 1.7 Hz, 1 H, H-(C4)); 5.16 (d, *J* = 2.8 Hz, 1 H, H-(C1′)); 4.26 (dd, *J* = 4.9, 2.8 Hz 1 H, H-(C2′)); 4.08 (m, 2 H, H-(C3′)[xlv]). ^{13}C NMR (151 MHz, H_2O/D_2O 9:1) δ_C 154.5 (C2); 135.3 (C5); 119.5 (C4); 75.8 (C1′); 62.5 (C2′).

General procedure A. Reaction of glycidaldehyde (69**) with 2-aminooxazole (**55**)**

2-Aminooxazole (**55**; 11.7 mg, 0.14 mmol) was dissolved in H_2O/D_2O (9:1, 1 mL) and glycidaldehyde (**69**; 10 mg, 0.128 mmol[xlvi]) was added. The solution was adjusted to the specified pH (3-12), continuously monitored and adjusted as required with 1M NaOH/HCl. ^{1}H NMR spectra were periodically acquired. In order to avoid more pH fluctuations, the solution was placed in a -20°C freezer overnight. Pentose aminooxazolines (**49**) and oxazolines **97** were identified by their characteristic H-C1' peaks.

pH	Time point (h)	^{1}H NMR integration relative to other discernible species integrated (%)			
		97	**49**	**69**	**12**
3	30	17	0	52	29
5	59	72	0	18	10
7	59	66	22	0	0
9	34	53	26	0	0
12	6	6	50	0	0

Table 9.11: Comparison of products at different pH ranges for the reaction described in General procedure A*, starting with glycidaldehyde (**69**; 128mM)/2-aminooxazole (**55**; 140mM). The percentage shown is relative to other discernible sugar species in solution including oxazoline **97**, aminooxazolines **49**, glyceraldehyde (**12**), and **69**. Results shown for time-point after which the system didn't show significant changes.*

[xlvi] The purity of **69** decreased over time to 90% (by ^{1}H NMR analysis with a standard), leading to the observed discrepancy between the weight added and the moles added.

Starting **69** concentration (mM)	Starting **55** concentration (mM)	T (°C)	Buffer	Time point (h)	**97** yield (%)
128	168	RT	-	14	68
128	168	40	-	14	55
63	70	RT	-	22	74
128	420	RT	-	12	66
384	140	RT	-	12	72
128	168	RT	Acetate[1]	20	68

*Table 9.12: Comparison of oxazoline **97** production under various conditions as described in G*eneral procedure A, *in order to improve yield of **97**. All reactions were done at pH 5. The percentage shown is based on ^{1}H NMR integration of H-C1' peaks and is relative to an internal standard .[1]0.5M acetate buffer. Results are shown for the time-points showing the best yield of **97**.*

General procedure B. Hydrolysis of oxazoline (97**) to aminooxazoline (**49**)**

Glycidaldehyde (**69**; 128mM), 2-aminooxazole (**55**; 1.1-1.3 eq) and KHP (24.2mM, internal standard) were dissolved in H_2O/D_2O (9:1). The solution was adjusted to pH 5 and then continuously monitored and adjusted as required with 1M HCl. ^{1}H NMR spectra were periodically acquired, showing conversion to oxazoline **97** (68%) after 13 h at RT. Oxazoline **97** solution was then adjusted to specified the pH (7–12) and incubated at RT. ^{1}H NMR spectra were again periodically acquired.

pH	Starting **55** concentration (mM)	Time point (h)	^{1}H NMR integration relative to other discernible species integrated (%)	
			97	**49**
7	140	32	84	12
9	140	25	67	26
12	168	20	0	70

Table 9.13: Comparison of products at different pH ranges for the reaction described in General procedure B. *Glycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; concentration shown above) were reacted at pH 5 for 13 h to form oxazoline **97**, after which the pH was raised to the value described above. The percentage shown is relative to other discernible sugar species in solution including oxazolines **97**, and aminooxazolines (**49**). Results shown for time-point after which the system didn't show significant changes.*

General procedure C. Reaction of glycidaldehyde (69**) with 2-aminooxazole (**55**) in the presence of metal cations**

2-Aminooxazole (**55**; 17.5 mg, 0.21 mmol) was dissolved in H_2O/D_2O (9:1, 1 mL) and the solution was adjusted to pH 5. MCl (0.42 mmol) was added, followed by addition of a solution of glycidaldehyde (**69**;15 mg, 0.19 mmol[xlvii]) in water (0.5 mL, pH 5). The solutions were immediately adjusted to pH 5 using a 100mM solution of the corresponding MOH. The pH was monitored and adjusted periodically and the solution was monitored by ^{1}H NMR. In order to avoid more pH fluctuations, the solution was placed in a -20°C freezer overnight. After 26 h at RT, the solutions were centrifuged and passed through a syringe filter. ^{1}H NMR spectra was acquired showing formation of oxazoline **97** in solution. To analyse the diastereomer selectivity, **97** was hydrolysed by raising pH to 12 using 1M NaOH. After 20 h at pH 12, the reactions were analysed by ^{1}H NMR.

To analyse the effect of salts in the reaction; the reaction was done with $CaCl_2$ (47 mg), $MgCl_2$ (40 mg), LiCl (18 mg) and NaCl (25 mg) respectively.

[xlvii] The purity of **69** decreased over time to 90% (by ^{1}H NMR analysis with a standard), leading to the observed discrepancy between the weight added and the moles added.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 5) (*partial assignment*)

Glyceraldehyde (**12**): 3.61 (m, 1 H, H^a-(C3)).

Glycidaldehyde (**69**): 3.12 (m, 1 H, H-(C2)).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

2-aminooxazole (**55**): 7.21 (s, 1 H, H-(C4)); 6.68 (s, 1 H, H-(C5)).

(2-Amino-4,5-dihydrooxazol-5-yl)((oxiran-2-yl)methanol (**97**):
97a: 5.62 (d, *J* = 2.6 Hz, 1 H, H-(C1′)).
97b: 5.60 (d, *J* = 2.7 Hz, 1 H, H-(C1′)).
97c: 5.57 (d, *J* = 2.8 Hz, 1 H, H-(C1′)).
97d: 5.55 (d, *J* = 2.6 Hz, 1 H, H-(C1′)).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 12) (*partial assignment*)

3-Hydroxypropanoic acid (**92**): 2.35 (t, *J* = 6.7 Hz, 1 H).

rac-Ribofuranosyl aminooxazoline (*ribo*-**49**): 5.70 (d, *J* = 5.1 Hz, 1 H, H-(C1′)).

rac-Arabinofuranosyl aminooxazoline (*arabino*-**49**): 5.81 (d, *J* = 5.5 Hz, 1 H, H-(C1′)).

rac-Xylofuranosyl aminooxazoline (*xylo*-**49**): 5.84 (d, *J* = 5.1 Hz, 1 H, H-(C1′)).

rac-Lyxofuranosyl aminooxazoline (*lyxo*-**49**f): 5.63 (d, *J* = 5.5 Hz, 1 H, H-(C1′)).

rac-Lyxopuranosyl aminooxazoline (*lyxo*-**49**p): 5.40 (d, *J* = 5.9 Hz, 1 H, H-(C1′)).

9.4.4. Glycidaldehyde (69), 2-aminooxazole (55) and phosphate experiments

General procedure D. Reaction of glycidaldehyde (69) with 2-aminooxazole (55) in the presence of P_i

2-Aminooxazole (**55**; 14 mg, 0.16 mmol) and glycidaldehyde (**69**; 10 mg, 0.128 mmol[xlviii]) were dissolved in H_2O/D_2O (9:1, 1 mL), followed by addition of sodium phosphate (mono- and/or di-basic; 0.14M-1.12M). The solution was adjusted to the specified pH (5-9), continuously monitored and adjusted as required with 1M NaOH/HCl. ^{1}H NMR spectra were periodically acquired. Aminooxazolines 5′-phosphate (**53**) and oxazolines **97** were identified by their characteristic H-(C1') peaks. **53** were confirmed by spiking with authentic samples.

[xlviii] The purity of **69** decreased over time to 90% (by ^{1}H NMR analysis with a standard), leading to the observed discrepancy between the weight added and the moles added.

pH	Starting **69** concentration (mM)	Starting **55** concentration (mM)	Phosphate (eq)	Time point (days)	^{1}H NMR integration relative to other discernible species integrated (%)		
					97	**53**[1]	**100**
5	128	168	6.4	11	0	54	4
7	128	140	1	14	9	62	15
7	128	140	3	14	7	60	18
7.4	128	168	3	10	0	48	30
7	128	168	6	6	0	46	39
7	13	17	3.5	7	0	46	8
9	128	160	6	4	0	49	26

Table 9.14: Comparison of products at different pH ranges and different amounts of P_i added, for the reaction described in General procedure D*. The percentage shown is relative to other discernible sugar species in solution including oxazoline **97**, aminooxazolines 5′-phosphate (**53**) and their open chain aromatic form **100**. [1]**53** H-C1′ signals overlap with aminooxazolines (**49**) signals. The yield therefore includes (minor) **49**.*

General procedure E. Phosphorylation of oxazoline (97**)** to **aminooxazolines 5′-phosphate (**53**)**

Glycidaldehyde (**69**; 128mM), 2-aminooxazole (**55**; 1.1-1.3 eq) and KHP (24.2mM, internal standard) were dissolved in H_2O/D_2O (9:1). The solution was adjusted to pH 5 and then continuously monitored and adjusted as required with 1M HCl. ^{1}H NMR spectra were periodically acquired, showing conversion to oxazoline **97** (68%) after 13 h at RT. A mixture of sodium dihydrogen phosphate and sodium hydrogen phosphate was added to the reaction in one portion (to give 0.14–1.12M phosphate, at the specified pH: 5-9). The solution pH was monitored, and no adjustments were required to maintain the desired pH. ^{1}H and ^{31}P NMR spectra were periodically acquired. Aminooxazolines 5′-phosphate (**53**) and oxazolines **97** were identified by their characteristic H-(C1') peaks. **53** were confirmed by spiking with authentic samples.

pH	Starting **55** concentration (mM)	Phosphate (eq)	Time point (days)	^{1}H NMR integration relative to other discernible species integrated (%)			
				97[1]	**53**[2]	**101**[3]	**100**
7	140	1	27	11	51	0	25
7	140	3	13	0	51	8	26
7	140	6	25	0	59	21	12
7[4]	168	8	19	0	58	7	13
9	140	1	19	0	72	0	9
9	140	3	13	0	61	12	9
9	168	8	4	0	59	3	12

Table 9.15: Comparison of products at different pH ranges and different amounts of phosphate added, for the reaction described in General procedure E. *Glycidaldehyde (**69**; 128mM) and 2-aminooxazole (**55**; concentration shown above) were reacted at pH 5 for 13 h to form oxazoline **97**, after which the sodium mono- and/or di-basic was added (equivalents shown above) and the pH was changed to the value described above. The percentage shown is relative to other discernible sugar species in solution including **97**, aminooxazolines 5′-phosphate (**53**), their open chain aromatic form **100**, and the open chain hydrate form **101**. [1]**97** H-C1′ signals might overlap with **55**·H_2O or **101** due to similarities in shift. They are only reported as such when the corresponding H-C5′ are seen. [2]**53** H-C1′ signals overlap with aminooxazolines **49** signals. The yield therefore includes (minor) **49**. [3]**101** H-C1′ signals might overlap with **55**·H_2O or **97** due to similarities in shift. They are only reported as such when the **97** H-C5′ aren't seen. [4]The reaction was done at pH 7 for 14 days after which the pH was raised to 9 for a further 5 days.*

General procedure F. Stereoisomer analysis of aminooxazolines 5′-phosphate (53**)**

The crude reaction mixture (1 mL, *as prepared in General Procedure D. or E.*) was brought to pH 10-11, and purified by ion exchange chromatography (Dowex® 1×8, 1 mL, OH^--form, prepared from Cl^--form and prewashed). The column was eluted as follows: water (3×5 mL), 0.5M aqueous HCO_2H (5 mL), and 1M aqueous HCO_2H (3×5 mL). The fractions containing aminooxazolines 5′-phosphate (**53**) were combined, diluted two-fold and lyophilised. The lyophilite was then dissolved in D_2O (0.5 mL) and ^{1}H and ^{31}P NMR spectra were acquired.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) for the CH_2O_2 fraction. (*partial assignment*)

(2*R*)-3-(2-Aminooxazole-5-yl)-2,3-dihydroxypropyl phosphate (**100**):
100b: 7.07 (s, 1 H, H-(C1′)).
100a: 7.05 (s, 1 H, H-(C1′)).

(2*S*)-2,3-dihydroxy-3-(4-hydroxy-2-iminooxazolidin-5-yl)propyl phosphate (**101**)[xlix]:
5.49 (d, *J* = 3.5 Hz, 1 H, H-(C1′)).
5.48 (d, *J* = 2.4 Hz, 1 H, H-(C1′)).
rac-Ribofuranosyl aminooxazoline (*ribo*-**53**): 5.94 (d, *J* = 5.1 Hz, 1 H, H-(C1′)).

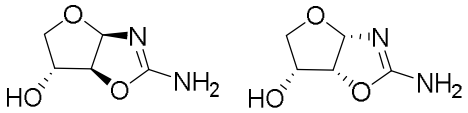
rac-Arabinofuranosyl aminooxazoline (*arabino*-**53**): 6.05 (d, *J* = 5.7 Hz, 1 H, H-(C1′)).

rac-Xylofuranosyl aminooxazoline (*xylo*-**53**): 6.07 (d, *J* = 5.3 Hz, 1 H, H-(C1′)).

[xlix] There are 8 overlapping isomers of **101**. Data is given for two that were assignable.

9.5. Experimental Procedure for 5. Multicomponent synthesis of 5′-aminooxazolines

*Tetrose aminooxazolines (**104**)*[354]



threo-**104** *erythro*-**104**

To a solution of glycolaldehyde (**3**; 14.4 mg, 0.24 mmol) in water (0.75 mL) was added 2-aminooxazole (**55**; 20.2 mg, 0.24 mmol). The solution was adjusted to pH 7 and heated at 40°C overnight. It was then lyophilised and the residue was dissolved in D_2O (82% crude yield) and used without further purification. HRMS (ESI^+) (*m/z*): $[M+H^+]^+$ calcd for formula $C_5H_9N_2O_3^+$, 145.0608; found, 145.0615.

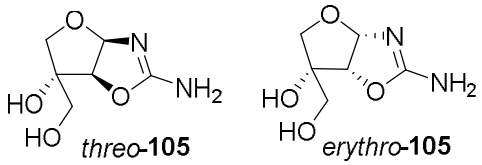
rac-Threose aminooxazoline (*threo*-**104**):

^1H NMR (600 MHz, D_2O) δ 5.84 (d, *J* = 5.1 Hz, 1 H, H-(C1′)); 4.78 (d, *J* = 5.1 Hz, 1 H, H-(C2′)); 4.32 (d, *J* = 2.4 Hz, 1 H, H-(C3′)); 3.79 (d, *J* = 10.9 Hz, 1 H, H^a-(C4′)); 3.56 (dd, *J* = 10.9, 2.4 Hz, 1 H, H^b-(C4′)). ^{13}C NMR (151 MHz, D_2O) δ_C 165.8 (C2); 98.5 (C1′); 87.3 (C2′); 74.2 (C3′); 70.4 (C4′).

rac-Erythrose aminooxazoline (*erythro*-**104**):

^1H NMR (600 MHz, D_2O) δ 5.71 (d, *J* = 5.1 Hz, 1 H, H-(C1′)); 4.83 (t, *J* = 5.1 Hz, 1 H, H-(C2′)); 4.27 (ddd, *J* = 9.5, 6.7, 5.1 Hz, 1 H, H-(C3′)); 3.90 (dd, *J* = 9.5, 6.7 Hz, 1 H, H^a-(C4′)); 3.20 (t, *J* = 9.5 Hz, 1 H, H^b-(C4′)). ^{13}C NMR (151 MHz, D_2O) δ_C 166.6 (C2); 98.9 (C1′); 81.9 (C2′); 71.1 (C3′); 66.1 (C4′).

*3′-Hydroxymethyl tetrose aminooxazolines (**105**)*[289,1]



threo-**105** *erythro*-**105**

[1] Prepared in the Powner lab by S. Islam.

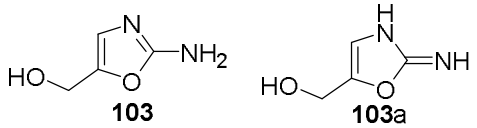
To a solution of 1,3 dihydroxyacetone (**8**; 450 mg, 5 mmol) in D_2O (5 mL) was added 2-aminooxazole (**55**; 840 mg, 5 mmol). The solution was adjusted to pD 7 and heated at 40°C overnight. It was then lyophilised and the residue was dissolved in D_2O (77% crude yield) and used without further purification. HRMS (ESI^+) (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_{11}N_2O_4^+$, 175.0713; found, 175.0720.

rac-3′-Hydroxymethyl threose aminooxazoline (*threo*-**105**):
^1H NMR (600 MHz, D_2O) δ 5.88 (d, *J* = 5.1 Hz, 1 H, H-(C1′)); 4.66 (d, *J* = 5.1 Hz, 1 H, H-(C2′)); 3.74 (m, 1 H, H^a-(C4″)), 3.70 (d, *J* = 10.5 Hz, 1 H, H^a-(C4′)); 3.62 (m, 1 H, H^b-(C4″)), 3.43 (d, *J* = 10.5 Hz, 1 H, H^b-(C4′)). ^{13}C NMR (151 MHz, D_2O) δ_C 165.6 (C2); 99.5 (C1′); 86.5 (C2′); 82.7 (C3′); 71.0 (C4′), 62.2 (C4″).

rac-3′-Hydroxymethyl erythrose aminooxazoline (*erythro*-**105**):
^1H NMR (600 MHz, D_2O) δ 5.74 (d, *J* = 5.2 Hz, 1 H, H-(C1′)); 4.58 (t, *J* = 5.2 Hz, 1 H, H-(C2′)); 3.75 (m, 1 H, H^a-(C4′)); 3.52 (m, 2 H, H-(C4″)); 3.32 (d, *J* = 9.6 Hz, 1 H, H^b-(C4′)). ^{13}C NMR (151 MHz, D_2O) δ_C 166.2 (C2); 99.2 (C1′); 82.7 (C2′); 80.4 (C3′); 67.5 (C4′), 64.0 (C4″).

Competition experiment of glycidaldehyde (69**), glyceraldehyde (**12**), and cyanamide (**4**)**

To a solution of glyceraldehyde (**12**; 12.5 mg, 140 μmol) and cyanamide (**4**; 5.8 mg, 140 μmol) in P_i buffer (0.84M, H_2O/D_2O 9:1, 1 mL, containing internal standard KHP) was added glycidaldehyde (**69**; 10 mg, 128 μmol[li]). The solution was adjusted to pH 7 and stirred at RT. The reaction was monitored by ^1H NMR periodically.



103 **103a**

[li] The purity of **69** decreased over time to 90% (by ^1H NMR analysis with a standard), leading to the observed discrepancy between the weight added and the moles added.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

(2-aminooxazol-5-yl)methanol: **103**a: 4.32 (s, 1H); **103**: 4.26 (s, 1H).

Glyceraldehyde (**12**): 3.60 (m, 1 H, H^a-(C3)).

Glycidaldehyde (**69**): 3.06 (m, 1 H, H-(C2)).

Reaction of glycidaldehyde (69**) and glycolaldehyde (**3**)**

To a solution of glycolaldehyde (**3**; 8.4 mg, 140 μmol) in P_i buffer (0.5M, H_2O/D_2O 9:1, 1 mL) was added glycidaldehyde (**69**; 10 mg, 140 μmol). The solution was adjusted to pH 7 and stirred at RT. The reaction was monitored by ^{1}H NMR periodically.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

Glycolaldehyde (**3**): 3.37 (d, J = 5.1 Hz, 2 H, H-(C2)).

Glycidaldehyde (**69**): 3.06 (m, 1 H, H-(C2)).

Reaction of glycidaldehyde (69**) and cyanamide (**4**)**

To a solution of cyanamide (**4**; 5.8 mg, 140 μmol) in P_i buffer (0.5M, H_2O/D_2O 9:1, 1 mL) was added glycidaldehyde (**69**; 10 mg, 140 μmol). The solution was adjusted to pH 7 and stirred at RT. The reaction was monitored by ^{1}H NMR periodically.

Reaction of glycolaldehyde (3**) and cyanamide (**4**)**

To a solution of cyanamide (**4**; 5.8 mg, 140 μmol) in P_i buffer (0.5M, H_2O/D_2O 9:1, 1 mL) was added glycolaldehyde (**3**; 8.4 mg, 140 μmol). The solution was adjusted to desired pH (5 or 7) and stirred at RT. The reaction was monitored by ^{1}H NMR periodically.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

2-aminooxazole (**55**): 7.15 (s, 1 H, H-(C4)); 6.63 (s, 1 H, H-(C5)).

2-aminooxazole hydrate (**55**·H_2O): 5.63 (dd, *J* = 6.1, 2.2 Hz, 1 H, H-(C4)); 4.65 (ABX, *J* = 10.3, 6.1 Hz, 1 H, H^a-(C5)); 4.38 (ABX, *J* = 10.3, 2.2 Hz, 1 H, H^b-(C5)).

Glycolaldehyde (**3**): 3.38 (d, *J* = 5.1 Hz, 2 H, H-(C2)).

Reaction of glyceraldehyde (12**), glycolaldehyde (**3**), cyanamide (**4**) and P_i**

A solution of glyceraldehyde (**12**; 125.2 mg, 1.4 mmol), glycolaldehyde (**3**; 100 mg, 1.68 mmol) and cyanamide (**4**; 70 mg, 1.68 mmol) in H_2O/D_2O (9:1, 5 mL) was prepared and P_i buffer (5 mL, 1M) was added. The solution was adjusted to pH 7, after which the solution was monitored by ^{1}H NMR periodically. After 6 days at RT, the solution was lyophilised and resuspended in D_2O (6 mL). The sample was centrifuged to remove excess non-solubilised inorganic phosphate. The supernatant was adjusted to pH 5 and analysed by ^{1}H NMR after 24 h. Products tetrose-aminooxazolines (**104**), 3′-hydroxymethyl tetrose aminooxazoline (**105**) and riboaminooxazoline (*ribo*-**49**) were confirmed by spiking with authentic samples (*Figure 5.9*).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 5) (*partial assignment*)

(2-aminooxazol-5-yl)methanol: **103**a: 4.39 (s, 1H); **103**: 4.30 (s, 1H).

rac-Threose aminooxazoline (*threo*-**104**): 6.07 (d, *J* = 5.1 Hz, 1 H, H-(C1′)); 5.25 (d, *J* = 5.1 Hz, 1 H, H-(C2′)).

rac-Erythrose aminooxazoline (*erythro*-**104**): 5.94 (d, *J* = 5.1 Hz, 1 H, H-(C1′)); 5.30 (t, *J* = 5.1 Hz, 1 H, H-(C2′)).

rac-3′-Hydroxymethyl threose aminooxazoline (*threo*-**105**): 6.01 (d, *J* = 5.0 Hz, 1 H, H-(C1′)).

rac-3′-Hydroxymethyl erythrose aminooxazoline (*erythro*-**105**): 5.98 (d, *J* = 5.0 Hz, 1 H, H-(C1′)).

rac-Ribofuranosyl aminooxazoline (*ribo*-**49**): 5.33 (t, *J* = 5.2 Hz, 1 H, H-(C2′)).

Reaction of glycidaldehyde (69**), glycolaldehyde (**3**) and cyanamide (**4**)**

To a solution of glycidaldehyde (**69**; 10 mg, 128 μmol[lii]) in H_2O/D_2O (9:1, 1 mL) was added glycolaldehyde (**3**; 10 mg, 168 μmol), and a solution of cyanamide (**4**; 7 mg, 168 μmol). The solution was adjusted to pH 7 and re-adjusted over time as needed. The reaction was monitored by ^{1}H NMR periodically.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

2-aminooxazole hydrate (**55**·H_2O): 5.65 (dd, *J* = 6.1, 2.1 Hz, 1 H, H-(C4)); 4.69 (ABX, *J* = 10.3, 6.1 Hz, 1 H, H^a-(C5)); 4.43 (ABX, *J* = 10.3, 2.1 Hz, 1 H, H^b-(C5)).

Glycolaldehyde (**3**): 4.96 (t, *J* = 5.2 Hz, 1 H, H-(C1)), 3.41 (d, *J* = 5.2 Hz, 2 H, H-(C2)).

Glycidaldehyde (**69**): 3.10 (dd, *J* = 4.1, 3.0 Hz, 1 H, H-(C2)).

General procedure for the reaction of glycidaldehyde (69**), glycolaldehyde (**3**), cyanamide (**4**), and P_i**

To a solution of glycidaldehyde (**69**; 10 mg, 128 μmol[lii]) in H_2O/D_2O (9:1, 0.5 mL) was added glycolaldehyde (**3**; 8.4 mg, 140 μmol), and a solution of cyanamide (**4**; 5.8 mg, 140 μmol) in P_i buffer (0.5 mL, 1M). The solution was adjusted to the desired pH (5-9) and monitored by ^{1}H NMR periodically.

[lii] The purity of **69** decreased over time to 90% (by ^{1}H NMR analysis with a standard), leading to the observed discrepancy between the weight added and the moles added.

Products tetrose-aminooxazolines (**104**), riboaminooxazoline 5'-phosphate (*ribo*-**53**), and xyloaminooxazoline (*xylo*-**49**) were confirmed by spiking with authentic samples (*Figures 5.10 and 5.11*).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

rac-Threose aminooxazoline (*threo*-**104**): 5.93 (d, J = 5.1 Hz, 1 H, H-(C1′)[liii]).

rac-Erythrose aminooxazoline (*erythro*-**104**): 5.80 (d, J = 5.1 Hz, 1 H, H-(C1′)[liii]); 5.02 (m, 1 H, H-(C2′)); 3.29 (t, J = 9.6 Hz, 1 H, H^b-(C4′)).

rac-Ribofuranosyl aminooxazoline 5′-phosphate (*ribo*-**53**): 5.80 (d, J = 5.1 Hz, 1 H, H-(C1′)[liii]; 5.10 (m, 1 H, H-(C2′)).

rac-Xylofuranosyl aminooxazoline (*xylo*-**49**): 5.93 (d, J = 5.1 Hz, 1 H, H-(C1′)[liii]); 5.02 (m, 1 H, H-(C2′)).

Epoxidation of acrolein (1**) in the presence of glycolaldehyde (**3**)**

To a solution of glycolaldehyde (**3**; 4.2 mg, 70 μmol) and hydrogen peroxide (7 μL of 10M aqueous solution, 70 μmol) in H_2O/D_2O (9:1, 0.5 mL, containing internal standard KHP) was added acrolein (**1**; 4.65 μL, 70 μmol). The solution was adjusted to pH 9 with 1M NaOH and monitored by ^{1}H NMR periodically.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 9) (*partial assignment*)

Glycolaldehyde (**3**): 3.48 (d, J = 5.2 Hz, 2 H, H-(C2)).

Glycidaldehyde (**69**): 3.17 (m, 1 H, H-(C2)).

[liii] The H-(C1′) peaks for *erythro*-**104** and *ribo*-**53** overlap, as well as for *threo*-**104** and *xylo*-**49**.

Epoxidation of acrolein (1**) in the presence of 2-aminooxazole (**55**)**

To a solution of 2-aminooxazole (**55**; 6 mg, 70 μmol) and hydrogen peroxide (7 μL of 10M aqueous solution, 70 μmol) in H_2O/D_2O (9:1, 0.5 mL, containing internal standard KHP) was added acrolein (**1**; 4.65 μL, 70 μmol). The solution was adjusted to pH 9 with 1M NaOH and monitored by ^{1}H NMR periodically.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 9) (*partial assignment*)

2-aminooxazole (**55**): 7.26 (s, 1 H, H-(C4)); 6.75 (s, 1 H, H-(C5)).

Glycidaldehyde (**69**) 3.17 (m, 1 H, H-(C2)).

Pentose aminooxazolines (**49**):
xylo-**49**: 5.90 (d, *J* = 5.1 Hz, 1 H, H-(C1′)).
arabino-**49**: 5.86 (d, *J* = 5.3 Hz, 1 H, H-(C1′)).
ribo-**49**: 5.76 (d, *J* = 5.0 Hz, 1 H, H-(C1′)).

9.6. Experimental Procedure for 6. From aminooxazolines 5′-phosphates to nucleotides

9.6.1. Anomerisation studies of aminooxazolines 5′-phosphates

General procedure G. Anomerisation of aminooxazolines 5′-phosphate (53**) in the absence of phosphate buffer**[liv]

Pentose aminooxazoline 5′-phosphate (**53**; 25 mg, 0.1 mmoles) was dissolved in H_2O/D_2O (1:1, 1 mL). The solution was adjusted to the desired pH (5 to 8) and incubated at the desired temperature (RT to 60°C). ^{1}H NMR were acquired periodically.

General procedure H. Anomerisation of aminooxazolines 5′-phosphate (53**) in the presence of phosphate buffer**

Pentose aminooxazoline 5′-phosphate (**53**; 25 mg, 0.1 mmoles) was dissolved in 1M Pi buffer (1 mL; H_2O/D_2O 1:1). The solution was adjusted to the desired pH (5 to 8) and incubated at the desired temperature (RT to 60°C). ^{1}H NMR were acquired periodically.

Products of the reaction were identified by spiking with authentic samples (oxazolidinones 5′-phosphate (**107**): *ribo*-**107**, *arabino*-**107** and *xylo*-**107**; arabinoaminooxazoline 5′-phosphate (*arabino*-**53**)) or were identified using their characteristic peaks (aromatic **100** and hydrate **101** were identified by comparison with the open chain aromatic and hydrate, respectively, of aminooxazolines (**49**): *ribo*-**49**, *arabino*-**49** and *xylo*-**49**[355]).

[liv] Reactions aimed at the synthesis of lyxose aminoooxazoline 5'-phosphate *lyxo*-**53** are described in the section "Reaction of D-lyxose-5-phosphate *lyxo*-**52** with cyanamide **4**", p.308.

^{1}H NMR (600 MHz, H_2O/D_2O 1:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment)*

D-Ribofuranosyl aminooxazoline 5′-phosphate (*ribo*-**53**): 5.85 (d, *J* = 5.1 Hz, 1 H, H-(C1′)); 5.21 (t, *J* = 5.1 Hz 1 H, H-(C2′)).

D-Arabinofuranosyl aminooxazoline 5′-phosphate (*arabino*-**53**): 5.92 (d, *J* = 5.6 Hz, 1 H, H-(C1′)); 5.16 (d, *J* = 5.6 Hz, 1 H, H-(C2′)).

D-Xylofuranosyl aminooxazoline 5′-phosphate (*xylo*-**53**): 6.04 (d, *J* = 5.1 Hz, 1 H, H-(C1′)); 5.19 (d, *J* = 5.2, 1 H, H-(C2′)).

D-Ribofuranosyl oxazolidinone 5′-phosphate (*ribo*-**107**): 5.70 (d, *J* = 5.2 Hz, 1 H, H-(C1′)); 5.03 (t, *J* = 5.2 Hz, 1 H, H-(C2′)).

D-Arabinofuranosyl oxazolidinone 5′-phosphate (*arabino*-**107**): 5.79 (d, *J* = 5.7 Hz, 1 H, H-(C1′)); 4.97 (d, *J* = 5.7 Hz, 1 H, H-(C2′)).

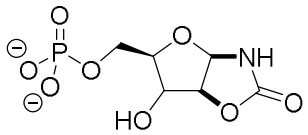
D-Xylofuranosyl oxazolidinone 5′-phosphate (*xylo*-**107**): 5.83 (d, *J* = 5.4 Hz, 1 H, H-(C1′)); 4.95 (d, *J* = 5.4 Hz, 1 H, H-(C2′)).

(2*R*, 3*S*)-3-(2-Aminooxazole-5-yl)-2,3-dihydroxypropyl phosphate (**100**a): 6.66 (s, 1 H, H-(C1′)).

(2*R*, 3*R*)-3-(2-Aminooxazole-5-yl)-2,3-dihydroxypropyl phosphate (**100**b): 6.63 (s, 1 H, H-(C1′)).

(2*S*)-2,3-Dihydroxy-3-(4-hydroxy-2-iminooxazolidin-5-yl)propyl phosphate (**101**):
ribo-**101**a: 5.46 (d, *J* = 4.2 Hz, 1H, H-(C1′)).
ribo-**101**b: 5.56 (d, *J* = 2.8 Hz, 1H, H-(C1′)).
arabino-**101**a: 5.47 (d, *J* = 4.2 Hz, 1H, H-(C1′)).
arabino-**101**b: 5.62 (d, *J* =2.8 Hz, 1H, H-(C1′)).
xylo-**101**a: 5.61 (d, *J* = 2.4 Hz, 1H, H-(C1′)).
xylo-**101**b: 5.87 (d, *J* = 4.2 Hz, 1H, H-(C1′)).
lyxo-**101**a: 5.72 (d, *J* = 2.4 Hz, 1H, H-(C1′)).
lyxo-**101**b: not observed.

General procedure I. Synthesis of oxazolidinones 5′-phosphates (107)[lv]



107

Furanoside 5′-phosphate (**52**; 0.15M) and sodium cyanate (2 eq) were dissolved in aqueous ammonium chloride (0.4M). The reaction mixture was stirred at 60°C for 7 h, cooled to RT and lyophilised. The white lyophilite was used directly without further purification.

*D-Ribose oxazolidinone 5′-phosphate (ribo-**107**)*

Starting from 176 mg *ribo*-**52**, yield=182 mg (95% crude yield). ^{1}H NMR (600 MHz, D_2O) δ 5.72 (d, *J* = 5.2 Hz, 1H, H-(C1')); 5.05 (dd, *J* = 5.5, 5.2 Hz, 1H, H-(C2')); 4.18 (dd, *J* = 8.6, 5.5 Hz, 1H, H-(C3')); 3.97 (ABXY, *J* = 12.1, 4.0, 2.1 Hz, 1H, H^a-(C5')); 3.91–3.79 (m, 2H, H-(C4'); H^a-(C5')). ^{13}C NMR (151 MHz, D_2O) δ 161.1 (C2); 86.0 (C1'); 80.5 (C2'); 78.2 (d, C4'); 70.7 (C3'); 62.5 (d, C5'). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled) δ 3.86. HRMS (ESI) (*m/z*): [M-H^-] calcd for formula $C_6H_9NO_8P$, 254.0071; found, 254.0074.

*D-Arabinose oxazolidinone 5′-phosphate (arabino-**107**)*

Starting from 91 mg *arabino*-**52**, yield = 95 mg (96% crude yield). ^{1}H NMR (600 MHz, D_2O) δ 5.79 (d, *J* = 5.8 Hz, 1H, H-(C1')); 4.98 (dd, *J* = 5.8, 1.2 Hz, 1H, H-(C2')); 4.40 (dd, *J* = 3.4, 1.2 Hz, 1H, H-(C3')); 4.12 (td, *J* = 6.3, 3.4 Hz, 1H, H-(C4')); 3.75–3.66 (m, 2H, H-(C5')). ^{13}C NMR (151 MHz, D_2O) δ 160.3 (C2); 87.2 (C2'); 87.1 (C1'); 85.0 (d, C4'); 75.6 (C3'); 64.1 (d, C5'). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled) δ 4.02. HRMS (ESI) (*m/z*): [M-H^-] calcd for formula $C_6H_9NO_8P$, 254.0071; found, 254.0079.

*D-Xylose oxazolidinone 5′-phosphate (xylo-**107**)*

Starting from 96 mg *xylo*-**52**, yield = 96 mg (91% crude yield). ^{1}H NMR (600 MHz, D_2O) δ 5.83 (d, *J* = 5.5 Hz, 1H, H-(C1')); 4.96 (d, *J* = 5.5 Hz, 1H, H-(C2'));

[lv] Prepared in the Powner lab by Dr. C. A. Fernández García.

4.41 (d, *J* = 2.6 Hz, 1H, H-(C3')); 4.16 (td, *J* = 6.3, 2.6 Hz, 1H, H-(C4')); 3.95 (ABXY, *J* = 11.0, 7.0, 6.3 Hz, 1H, H^a-(C5')); 3.87 (ABXY, *J* = 11.0, 7.0, 6.3 Hz, 1H, H^a-(C5')).^{13}C NMR (151 MHz, D_2O) δ 160.7 (C2); 86.8 (C1'); 85.7 (C2'); 79.3 (d, C4'); 73.6 (C3'); 61.3 (d, C5'). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled) δ 4.73. HRMS (ESI) (*m/z*): [M+Na^+] calcd for formula $C_6H_{10}NNaO_8P$, 278.0036; found, 278.0034.

Tables showing the results of the anomerisation studies of arabinoaminooxazoline 5'-phosphate (*arabino*-**53**) at RT. Comparison of pH and presence of P_i buffer on anomerisation rate and formation of aromatic species **100**b, hydrates *ribo*-/*arabino*-**101**, and oxazolidinones 5'-phosphate *ribo*-**107** and *arabino*-**107**. The percentage of each compound in solution is based on ^{1}H NMR integration relative to other species in solution.

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	96.2	0	0	0.1	3.8
1	0	74.8	0.2	0	0.7	24.3
2	0	72.9	1.2	0	1.2	24.8
3	0	73.5	0.8	0	1.3	24.3
5	0.7	71.4	1	0	2.4	24.6
7	2	70.7	0.9	0	2.8	23.6
9	3.8	67.2	1.7	0.3	3	24
10	5	66.7	1.6	0.5	3.1	23.1
14	6.3	67.3	1	0.3	3.3	21.8
21	10.6	62.9	0.2	0.6	2.6	23.1

*Table 9.16: Time course reaction of arabino-**53** (0.1M) at RT in H_2O/D_2O (1:1) at pH 5.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	95.8	0	0	0.2	4
1	0	68.3	1.8	0	2.9	26.9
2	1	64.8	1.7	0	5.8	26.7
3	2.6	62.5	1.7	0	8.5	24.8
5	6.1	57.6	1.6	0	10.8	23.8
7	10.4	55	0.8	0	12	21.9
9	12.6	53.9	0.7	0	11.6	21.3
10	16.9	51.7	0.3	0	11.9	19.1
14	25.8	44.9	0.2	0.7	11.9	16.6
21	35.3	37.2	0.3	1.2	10	16.1

*Table 9.17: Time course reaction of arabino-**53** (0.1M) at RT in H_2O/D_2O (1:1) at pH 6.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	96.4	0	0	0.3	3.3
1	1.4	65	1.6	0	6.6	25.3
2	1.1	62.9	1.1	0	11.5	23.4
3	3.8	55.7	1.4	0.5	16.2	22.5
5	5.1	50.5	0.8	0.1	23.2	20.5
7	10.4	42.7	1.2	1	26.3	18.5
9	9.7	42.6	0.3	0.5	29.4	17.5
10	15.8	36.9	0.4	0.7	31	15.2
14	24.4	30.2	0.2	1	32	12.2
21	25.3	29.3	0.1	1	32.3	12.1

*Table 9.18: Time course reaction of arabino-**53** (0.1M) at RT in H_2O/D_2O (1:1) at pH 7.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	95.6	0	0	0.1	4.3
1	0	72.6	0.8	0	1.1	25.5
2	1.6	67.4	2	0	2	27.1
3	3.4	66.5	1.9	0.2	2.5	25.7
5	6.7	65.3	1.1	0	2.6	24.2
7	10.3	63.4	0.6	0	2.5	23.2
9	14.9	57.8	1.6	1.2	2.3	22.3
10	17.9	55.4	1.5	1.5	2.7	21
14	25.7	53.2	0.1	0.7	1.3	19.1
21	32.5	43.7	1.2	3.1	2.1	17.3

*Table 9.19: Time course reaction of arabino-**53** (0.1M) at RT in 1M P_i buffer at pH 5.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	94.4	0	0	0.44	5.2
1	1	63.6	1.9	0	3.5	30.1
2	3.4	59	2.1	0.4	5.1	30
3	6.3	58.1	1.2	0.2	6.1	28.2
5	15	50	1.7	1.4	7.4	24.6
7	21.6	46	1.6	1.8	7	22.1
9	28.1	41.2	1.4	2.2	6.7	20.4
10	33	38.3	1.3	2.7	6.8	17.9
14	43.8	30.6	1	3.9	6.2	14.5

*Table 9.20: Time course reaction of arabino-**53** (0.1M) at RT in 1M P_i buffer at pH 6.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	92.9	0	0	0.5	6.6
1	1.9	56.2	1.7	0.6	9.2	30.5
2	3.5	50.3	1.7	0.8	13.7	30.2
3	6.3	46.3	1.5	1.3	18.6	26
5	13.3	38	1.1	2.1	23.8	21.7
7	19.7	33	1	3	24.9	18.4
9	26.7	27.8	0.8	3.5	25.8	15.4
10	31.1	25.2	0.5	4	26	13.2
14	40.5	18.4	2.3	5.3	23.9	9.6
21	48.3	13.9	3.6	6.1	20.6	7.6

*Table 9.21: Time course reaction of arabino-**53** (0.1M) at RT in 1M P_i buffer at pH 7.*

Tables showing the results of the anomerisation studies of arabinoaminooxazoline 5'-phosphate (*arabino*-**53**) at 40°C. Comparison of pH and presence of P_i buffer on anomerisation rate and formation of aromatic species **100**b, hydrates *ribo*-/*arabino*-**101**, and oxazolidinones 5'-phosphate *ribo*-**107** and *arabino*-**107**. The percentage of each compound in solution is based on ^{1}H NMR integration relative to other species in solution.

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	92	8	0	0	0
1	2.5	72.4	1.7	0	3.3	20.3
2	5.6	68.7	1.6	1.1	3.6	19.3
3	7.7	66.9	1.6	1.5	3.3	19.
5	13.2	61.4	1.7	2.5	3.4	17.7
7	17	58.3	1.6	3.4	3.0	16.7
10	21.6	54.4	14	4.9	2.5	15.2
14	26.1	49.5	1.4	6.7	2.4	13.9

*Table 9.22: Time course reaction of arabino-**53** (0.1M) at 40°C in H_2O/D_2O (1:1) at pH 5.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	90.6	0	0	0	9.4
1	13.9	46.3	1.5	1.2	22.1	15.1
2	27.3	36.1	1.1	1.6	22.3	11.6
3	36.6	29.4	1	2.5	20.4	10.1
5	49.4	19.9	1.8	3.7	18.5	6.6
7	53.3	16.6	3.5	4.8	16.1	5.7
10	57	13.3	4.7	6.1	14.3	4.6
14	56.5	11.6	7	8	12.8	4.2

*Table 9.23: Time course reaction of arabino-**53** (0.1M) at 40°C in H_2O/D_2O (1:1) at pH 6.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	97.2	0	0	0.2	2.6
1	10.6	38.2	1.2	1.6	35.8	12.6
2	23.2	25.6	0.8	2.2	39.7	8.5
3	31.1	20	1.2	2.8	38	6.9
5	41.8	13.1	2.3	4	34.4	4.5
7	45.1	10.8	3.4	5	32	3.7
9	46.9	9.7	4.6	5.6	30	3.2
10	46.2	9.1	5.5	6.2	29.4	3.7
14	45.5	8.5	7.6	7.9	27.4	3.2

*Table 9.24: Time course reaction of arabino-**53** (0.1M) at 40°C in H_2O/D_2O (1:1) at pH 7.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	89.1	0.6	0	0.8	9.5
1	15.9	57.3	1.8	1.7	5.6	17.7
2	29.4	46.9	1.4	2.9	5	14.4
3	39.3	39.5	1.3	3.8	4.2	12
5	51.6	27.5	2.8	5.8	3.9	8.5
7	58.1	21	4.2	7.1	3.4	6.3
9	60.6	15.7	6.9	8.9	3	4.9
10	59.2	12.8	10.9	10.9	2.2	4.1
14	0	89.1	0.6	0	0.8	9.5

*Table 9.25: Time course reaction of arabino-**53** (0.1M) at 40°C in 1M P_i buffer at pH 5.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	86.8	0.9	0	1.7	10.7
1	25	41.6	1.1	2.6	14.1	15.6
2	42.5	28.8	1	4.2	12.8	10.7
3	52.2	21.3	1	5.3	11.9	8.3
5	57.9	13.6	6.6	6.9	9.4	5.5
7	57.8	11.3	9.8	8.4	8.2	4.6
9	54.8	9.3	15	10.2	6.8	3.9
10	49.3	8.3	21.4	12.1	5.6	3.3
14	0	86.8	0.9	0	1.7	10.7

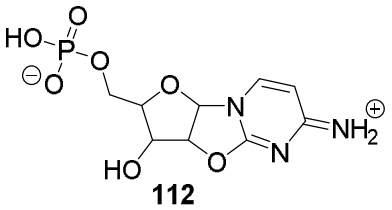
*Table 9.26: Time course reaction of arabino-**53** (0.1M) at 40°C in 1M P_i buffer at pH 6.*

Time (days)	^{1}H NMR integration relative to other species integrated (%)					
	ribo-**53**	*arabino*-**53**	*ribo*-**107**	*arabino*-**107**	**100**b	**101**b
0	0	85.6	0.9	0	2.2	11.3
1	17.9	29.8	0.9	3.3	35.1	13
2	32.6	17.3	1.9	4.8	35.9	7.5
3	40.1	12.8	3.3	6.0	32.6	5.2
5	44.6	8.2	7.1	7.6	28.9	3.6
7	44	6.9	11.3	9.1	25.6	3.3
9	41.8	6.3	15.9	10.2	22.7	3.1
10	40.9	6.1	17.6	10.5	22	2.9
14	36.5	5.6	25.5	12.3	17.9	2.2

*Table 9.27: Time course reaction of arabino-**53** (0.1M) at 40°C in 1M P_i buffer at pH 7.*

9.6.2. Phosphoryl transfer

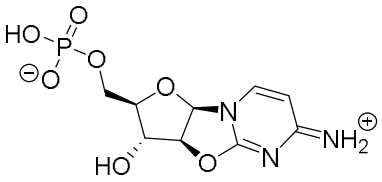
General procedure for the preparative synthesis of [1′,2′-cis]-2,2′-anhydrocytidine-5'-phosphate hydrochloride salts (112**)**[33]



112

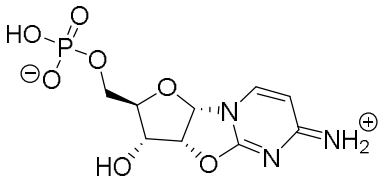
Aminooxazoline 5′-phosphate (**53**; 1.07 g, 4.21 mmoles) and sodium hydrogen phosphate (525.5 mg, 4.38 mmoles, 1.1 eq) were dissolved in water (50 mL), and the solution was adjusted to pH 6.5 with 4M NaOH. Cyanoacetylene (**6**; 35 mL, 0.85M in H_2O, 7 eq) was added and the solution was readjusted to pH 6.5 using 1M NaOH/HCl, and stirred at RT for 16 h. The solution was lyophilised and the lyophilite purified *via* ion exchange chromatography with Dowex® 50W×8 resin (H^+-form, prewashed), eluting with water (150 mL), then 0.125M HCl (150 mL), 0.25M HCl (150 mL), 0.5M HCl (150 mL), 1M HCl (150 mL) and 2M HCl (150 mL). The fractions containing the desired product (0.5M-1M) were concentrated and lyophilised to afford D-furanosyl-2,2′-anhydrocytidine 5′-phosphate (**112**) as white powder. (*arabino*-**112**: 95%; *ribo*-**112**: 80% starting with the corresponding aminooxazoline 5′-phosphate (**53**)).

*D-Arabinofuranosyl-2,2'-anhydrocytidine 5'-phosphate (arabino-**112**)*



^{1}H NMR (600 MHz, D_2O) δ 8.02 (d, *J* = 7.4 Hz, 1 H, H-(C6)); 6.57 (d, *J* = 6.2 Hz, 1 H, H-(C1′)); 6.50 (d, *J* = 7.4 Hz, 1 H, H-(C5)); 5.51 (d, *J* = 6.2 Hz, 1 H, H-(C2′)); 4.78 (1H, H-(C3′), *obscured by residual solvent peak*); 4.49 (t, *J* = 2.4 Hz, 1 H, H-(C4′)); 4.02-3.95 (m, 2 H, H-(C5′)). ^{13}C NMR (150 MHz, D_2O) δ_C 168.3 (C4); 160.6 (C2); 140.6 (C6); 103.7 (C5); 92.3 (C1′); 91.2 (C2′); 89.2 (C4′); 76.2 (C3′); 65.3 (d, C5′). ^{31}P NMR (121 MHz, D_2O, ^{1}H-decoupled): δ 2.36. HRMS (ES^-) (*m/z*): $[M]^-$ calcd for formula $C_9H_{13}N_3O_7P^+$, 306.0486; found, 306.0483

*D-Ribofuranosyl-2,2'-anhydrocytidine 5'-phosphate (ribo-**112**)*[33,lvi]



^{1}H NMR (600 MHz, D_2O) δ 8.19 (d, *J* = 7.4 Hz, 1H, H-(C6)); 6.69 (d, *J* = 7.4 Hz, 1H, H-(C5)); 6.62 (d, *J* = 5.4 Hz, 1H, H-(C1′)); 5.68 (dd, *J* = 5.6, 5.4 Hz, 1H, H-(C2′)); 4.58 (dd, *J* = 8.7, 5.6 Hz, 1H, H-(C3′)); 4.28 (ABXY, *J* = 11.0, 5.6, 2.6 Hz, 1H, H^a-(C5′)); 4.17-4.07 (m, 2H, H-(C4′), H^b-(C5)). ^{13}C NMR (151 MHz, D_2O) δ 168.2 (C4); 161.6 (C2); 141.0 (C6); 103.9 (C5); 90.9 (C1′); 84.1 (C2′); 80.1 (d, C4′); 70.1 (C3′); 63.4 (d, C5′). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled) δ 0.49.

General procedure for the dry-state phosphorylation of anhydrocytidine 5'-phosphate (112**)**

Protocol A: Anhydrocytidine 5'-phosphate (**112**; 0.12 mmol) was dissolved in H_2O (0.5 mL), along with ammonium chloride (2 eq), urea (**28**; 5-10 eq), and ammonium dihydrogen phosphate (1 eq) or pyrophosphate (0.5 eq). The resulting solution was applied evenly to glass fibre discs (24mm, 0.1 mL per disc) and dried at 40°C for 48 h. The discs were then heated at 100°C for 24 h, then washed with H_2O and centrifuged (3×5mL). The combined supernatant was lyophilised, dissolved in D_2O and centrifuged to remove particulates (3×1mL). The combined supernatant was lyophilised, dissolved in D_2O and analysed by ^{1}H and ^{31}P NMR.

Protocol B: Anhydrocytidine 5'-phosphate (**112**; 0.12 mmol) was dissolved in H_2O (0.5 mL), along with ammonium chloride (2 eq), urea (**28**; 5-10 eq). The resulting solution was applied evenly to glass fibre discs (24mm, 0.1 mL per disc) and dried at 40°C for 48 h. The discs were then heated at 100°C for 24 h, then washed with H_2O and centrifuged (3×5mL). The combined supernatant was lyophilised, dissolved in D_2O and centrifuged to remove particulates (3×1mL). The combined supernatant was lyophilised, dissolved in D_2O and analysed by ^{1}H and ^{31}P NMR.

[lvi] Prepared in the Powner lab by Dr. C. A. Fernández García.

Major products of D*-arabinofuranosyl-2,2′-anhydrocytidine 5′-phosphate (arabino-**112**) phosphorylations*

^{1}H NMR (400 MHz, D_2O) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment)*

D-Arabinofuranosyl-2,2′-anhydrocytidine 5′-phosphate (*arabino*-**112**): 5.72 (d, *J* = 6.1 Hz 1 H, H-(C2′)). ^{31}P NMR (121 MHz, D_2O, ^{1}H-decoupled): δ 3.54

β-D-Ribofuranosyl cytidine-2′,3′-cyclic-5′-bisphosphate (β-**44**): 5.26-5.22 (m, 1H, H-(C2′)); 5.19-5.14 (m, 1H, H-(C3′)); 4.33-4.28 (ABXY, 1H, H^a-(C5′)); 4.24 (ABXY, *J* = 12.4, 5.7 Hz, 1H, H^b-(C5′)). ^{31}P NMR (121 MHz, D_2O, ^{1}H-decoupled): δ 22.72, 3.34.

β-D-Ribofuranosyl cytidine-2′,3′-cyclic phosphate (β-**43**): 5.28 (dt, *J* = 6.4, 2.8 Hz, 1H, H-(C2′)); 5.08 (dt, *J* = 12.1, 6.4 Hz, 1H, H-(C3′)); 4.05 (ABX, *J* = 12.4, 3.6 Hz, 1H, H^a-(C5′)); 3.97 (ABX, *J* = 12.4, 5.7 Hz, 1H, H^b-(C5′)). ^{31}P NMR (121 MHz, D_2O, ^{1}H-decoupled): δ 23.11.

Arabino-**112** and β-**43** were confirmed by spiking with authentic samples. β-**44** was inferred based on similarities with β-**43** and the presence of additional coupling patterns, ^{31}P-^{1}H HMBC correlation between H_2-(C5′) and a phosphorous signal and in accordance with literature data[33].

Major products of D*-ribofuranosyl-2,2′-anhydrocytidine 5′-phosphate (ribo-**112**) phosphorylations*[lvii]

^{1}H NMR (400 MHz, D_2O) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment)*

D-Ribofuranosyl-2,2′-anhydrocytidine 5′-phosphate (*ribo*-**112**): 5.72 (d, *J* = 5.1 Hz, 1 H, H-(C2′)). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled): δ 1.29.

α-D-Ribofuranosyl cytidine-2′,3′-cyclic-5′-bisphosphate (α-**44**): 5.27 (dt, *J* = 10.3, 5.4 Hz, 1H, H-(C2′)); 5.14 (dt, *J* = 10.9, 5.4 Hz, 1H, H-(C3′)); 4.27-4.00 (2H, H-(C5′)). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled): δ 19.38, 1.69.

Ribo-**112** was confirmed by spiking with an authentic sample. α-**44** was inferred based on similarities with α-**43** and the presence of additional coupling patterns, ^{31}P-^{1}H HMBC correlation between H_2-(C5′) and a phosphorous signal and in accordance with literature data[33].

General procedure for phosphoryl-transfer in anhydrocytidine 5'-phosphate (112**)**

Anhydrocytidine 5'-phosphate (**112**; 0.12 mmol) was dissolved in H_2O (0.5 mL), along with ammonium chloride (2 eq), and urea (**28**; 5-20 eq). The resulting solution was applied evenly to glass fibre discs (24mm, 0.1 mL per disc) and dried at 40°C for 48 h. The discs were then heated at 100°C for 24 h, then washed with H_2O and centrifuged (3×5mL). The combined supernatant was lyophilised, dissolved in D_2O and centrifuged to remove particulates (3×1mL). The combined supernatant was lyophilised, dissolved in D_2O and analysed by ^{1}H and ^{31}P NMR.

[lvii] *Ribo*-**115** dry-state phosphorylation and phosphoryl-transfer experiments were performed in the Powner lab by Dr. C. A. Fernández García.

*Major products of D-arabinofuranosyl-2,2′-anhydrocytidine 5′-phosphate (arabino-**112**) phosphoryl-transfer experiments*

^{1}H NMR (400 MHz, D_2O) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment)*

D-Arabinofuranosyl-2,2′-anhydrocytidine 5′-phosphate (*arabino*-**112**): 5.62 (d, *J* = 6.0 Hz 1 H, H-(C2′)). ^{31}P NMR (121 MHz, D_2O, ^{1}H-decoupled): δ 0.18.

β-D-Ribofuranosyl cytidine-2′,3′-cyclic-5′-bisphosphate (β-**44**): 5.17-5.02 (m, 2H, H-(C2′)/H-(C3′), *partially obscured by H-(C2')/H-(C3') of β-**43***); 4.26-4.12 (2H, H-(C5′)). ^{31}P NMR (121 MHz, D_2O, ^{1}H-decoupled): δ 20.01, 0.61.

β-D-Ribofuranosyl cytidine-2′,3′-cyclic phosphate (β-**43**): 5.18 (dt, *J* = 6.8, 3.0 Hz, 1H, H-(C2′)); 4.98 (1H, H-(C3′), *partially obscured by H-(C3') of β-**44***); 3.95 (ABX, *J* = 12.3, 3.5 Hz, 1H, H^a-(C5′)); 3.87 (ABX, *J* = 12.3, 5.7 Hz, 1H, H^b-(C5′)). ^{31}P NMR (121 MHz, D_2O, ^{1}H-decoupled): δ 20.42.

Arabino-**112** and β-**43** were confirmed by spiking with authentic samples. β-**44** was inferred based on similarities with β-**43** and the presence of additional coupling patterns, ^{31}P-^{1}H HMBC correlation between H_2-(C5′) and a phosphorous signal and in accordance with literature data[33].

Major products of D*-ribofuranosyl-2,2′-anhydrocytidine 5′-phosphate (ribo-**112**) phosphoryl-transfer experiments*[lviii]

^{1}H NMR (400 MHz, D_2O) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment)*

D-Ribofuranosyl-2,2′-anhydrocytidine 5′-phosphate (*ribo*-**112**): 5.65 (d, *J* = 5.5 Hz, 1 H, H-(C2′)). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled): δ 0.51.

α-D-Ribofuranosyl cytidine-2′,3′-cyclic-5′-bisphosphate (α-**44**): 5.28-5.08 (2H, H-(C2′)/H-(C3′)). ^{31}P NMR (162 MHz, D_2O, ^{1}H-decoupled): δ 19.37, 0.85.

Ribo-**112** was confirmed by spiking with an authentic sample. α-**44** was inferred based on similarities with α-**43** and the presence of additional coupling patterns, ^{31}P-^{1}H HMBC correlation between H_2-(C5′) and a phosphorous signal, and in accordance with literature data[33].

[lviii] *Ribo*-**115** dry-state phosphorylation and phosphoryl-transfer experiments were performed in the Powner lab by Dr. C. A. Fernández García.

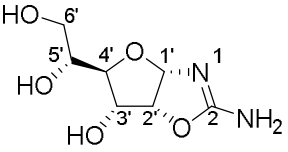
9.7. Experimental Procedure for 7. Purine synthesis

9.7.1. Synthesis of 5′-oxoriboaminooxazoline

General procedure for the time course reaction of D-allose (121**) and cyanamide (**4**)**

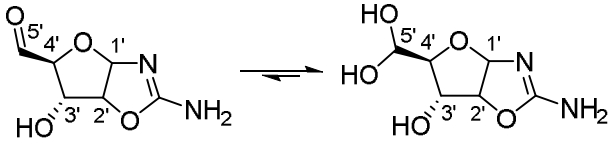
To a solution of D-allose (**121**; 19.5 mg, 0.13 mmol) in aqueous ammonia (3.5%, 1 mL) was added cyanamide (**4**; 2-4 eq). The resulting solution was stirred at the desired temperature (RT-80°C) and the reaction progress was monitored by ^{1}H NMR analysis. Heating at 80°C for 1 h (**4**: 4 eq) afforded 87% crude D-alloaminoxazoline (**122**; based on ^{1}H NMR integration of species in solution) and was chosen as the bases for scale-up to gram scale production.

*D-Allofuranosyl aminooxazoline (**122**)*



To a solution of D-allose (**121**; 1.0 g, 5.6 mmol) in aqueous ammonia (3.5%, 10 mL) was added cyanamide (**4**; 0.93 g, 22 mmol). The resulting solution was stirred at 80°C for 1 h. The reaction was cooled to 4ºC and after 16 h white crystals were collected by filtration, washed with ice-cold ethyl acetate and dried under vacuum to yield D-allofuranosyl aminooxazoline (**122**) as a fine white powder (645 mg, 58%). M.p. >173°C (decomposes). $[\alpha]_D^{25}$= +10.16 (c=0.5, water). IR (cm^{-1}) 3380 (N-H), 3330-3170 (O-H), 3060-2880 (C-H); 1666 (C=N). ^{1}H NMR (600 MHz, D_2O) δ 5.68 (d, *J* = 5.2 Hz, 1 H, H-(C1′)); 4.87 (t, *J* = 5.2 Hz, 1 H, H-(C2′)); 4.17 (dd, *J* = 9.3, 5.2 Hz, 1 H, H-(C3′)); 3.91 (dt, *J* = 7.2, 3.5 Hz, 1 H, H-(C5′)); 3.62 (ABX, *J* = 11.9, 3.5 Hz, 1 H, H^{a}-(C6′)); 3.56 (ABX, *J* = 11.9, 7.2 Hz, 1 H, H^{b}-(C6′)); 3.48 (dd, *J* = 9.3, 3.5 Hz, 1 H, H-(C4′)). ^{13}C NMR (151 MHz, H_2O+D_2O) δ_C 166.7 (C2); 97.6 (C1′); 82.9 (C2′); 77.6 (C4′); 71.5 (C5′); 71.2 (C3′); 63.1 (C6′). HRMS (*m/z*): [M+Na^{+}]$^{+}$ calcd for formula $C_7H_{12}N_2O_5Na^{+}$, 227.0638; found, 227.0623.

*5′-Oxo-D-ribofuranosyl aminooxazoline (**123**)*



*Method 1 – through D-alloaminooxazoline (**122**):*

A modified protocol for oxidative diol-cleavage based around a published protocol by Paquette and co-worker[337] was used to synthesis 5′-oxo-D-ribofuranosyl aminooxazoline (**123**). To a solution of D-alloaminooxazoline (**122**; 217 mg, 1.08 mmol) in water (7.2 mL) was added sodium metaperiodate (346 mg, 1.62 mmol). The resulting solution was stirred at RT for 2 h. The reaction was then cooled in an ice-water bath, and barium acetate (0.21 g, 0.81 mmol) was added to precipitate sodium iodate. After 10 min the mixture was centrifuged and the supernatant was collected. The resulting pellet was then washed with water (3×5 mL). The combined supernatant was concentrated *in vacuo* and the residue was dissolved in a minimum volume of hot water. The sample was dry-loaded onto silica (2 mL) and the powder was subsequently loaded onto a small silica plug and eluted with MeOH:EtOAc (30:70, 200 mL). The desalted solution was dry-loaded onto silica and the sample was concentrated to a fine free-following (white) powder. The sample was then purified by FCC (5:93:2 to 30:68:2, MeOH:EtOAc:Et_3N). Fractions containing 5′-oxo-D-ribofuranosyl aminooxazoline (**123**) were combined and reduced to a third of the volume *in vacuo* and the volume was made up to 20 mL with water. The solution was adjusted to pH 9 with OH^- Dowex® 1×8 resin column (prepared from Cl^--form and prewashed) and the solution was reduced to 7 mL *in vacuo* to remove Et_3N. Dilution and pH adjustment was repeated until the pH remained stable after evaporation and NMR analysis confirmed that there was no residual Et_3N in solution. The resultant solution was then filtered and lyophilised. 5′-Oxo-D-ribofuranosyl aminooxazoline (**123**) was obtained as a fine white powder (68 mg, 36%). TLC dip tests with 2,4 DNP (with positive and negative controls, glycolaldehyde (**3**) and riboaminooxazoline (*ribo*-**49**) respectively) showed presence of expected carbonyl group in our product. M.p. >170°C

(decomposes). $[\alpha]_D^{25}$= -0.002 (c=0.02, water). IR (cm^{-1}) 3214 (O-H), 1648 (C=N). ^{1}H NMR (600 MHz, D_2O) δ 5.87 (d, *J* = 5.2 Hz, 1 H, H-(C1′)); 5.03 (d, *J* = 4.1 Hz, 1 H, H-(C5′)); 4.91 (t, *J* = 5.2 Hz, 1 H, H-(C2′)); 4.11 (dd, *J* = 9.0, 5.2 Hz, 1 H, H-(C3′)); 3.45 (dd, *J* = 9.0, 4.1 Hz, 1 H, H-(C4′)). ^{13}C NMR (151 MHz, D_2O) δ_C 166.5 (C2); 97.7 (C1′); 89.9 (C5′); 82.9 (C2′); 79.6 (C4′); 72.4 (C3′). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_6H_9N_2O_4^+$, 173.0557; found, 173.0546.

*Method 2 - through oxidation of riboaminooxazoline (ribo-**49**):*

Following Angelin *et al.*'s general procedure for oxidation[338]: Sodium bicarbonate (1.3 g, 16 mmol) and TEMPO (2 mg, 13 μmol) were added to a stirred solution of D-ribofuranosyl aminooxazoline (*ribo*-**49**; 90 mg, 0.52 mmol) in DMF (100 mL). The reaction was cooled to 0°C and TCC (90 mg, 0.4 mmol) was added. After 7 h of continuous stirring at 0°C, the reaction mixture was filtered, and the solvent removed *in vacuo*. The residue was dissolved in D_2O and analysed by ^{1}H NMR, showing incomplete reaction (20% conversion by ^{1}H NMR integration).

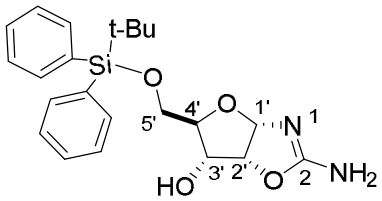
^{1}H NMR (600 MHz, D_2O) δ ppm (*partial assignment*)

D-Ribofuranosyl aminooxazoline (*ribo*-**49**): 5.79 (H-(C1′), *overlap with H-(C1') of **123**)*); 5.16 (H-(C2′), *overlap with H-(C2') of **123**)*); 4.09 (dd, *J* = 9.5, 5.2 Hz, 1 H, H-(C3′)); 3.85 (ABX, *J* = 12.8, 2.2 Hz, 1 H, H^a-(C5′)); 3.77 (ddd, *J* = 9.5, 4.7, 2.2 Hz, 1 H, H-(C4′)), 3.65 (ABX, *J* = 12.8, 4.7 Hz, 1 H, H^b-(C5′)).

5′-Oxo-D-ribofuranosyl aminooxazoline (**123**): 5.80 (H-(C1′), *overlap with H-(C1') of ribo-**49**)*); 5.06 (d, *J* = 4.0 Hz, 2 H, H-(C5′)); 4.20-4.15 (m, 1 H, H-(C3′)); 3.70-3.65 (m, 1 H, H-(C4′))[lix].

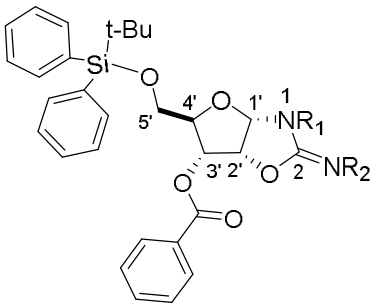
[lix] The remaining proton signal for **123**, H-(C2′), overlaps with the stronger H-(C2′) *ribo*-**49** signal.

*5′-tert-Butyldiphenylsilyl-ether-D-ribofuranosyl aminooxazoline (**130**)*



To a stirred solution of D-riboaminooxazoline (*ribo*-**49**; 1.5 g, 8.6 mmol) in pyridine (10 mL) was added DMAP (catalytic amount) and TBDPS-Cl (2.5 ml, 9.5 mmol). The resulting solution was stirred under nitrogen at RT for 20 h. The solvent was removed *in vacuo* after quenching with methanol (2 mL) and the solution was analysed by ^{1}H NMR in MeOD. This showed near quantitative conversion of *ribo*-**49** to 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl aminooxazoline (**130**) with only 5% impurities, so the bulk of the reaction was carried on without purification. A small portion of the compound (204 mg) was purified further for characterisation by dissolving the residue in a minimal amount of methanol (containing 0.1% NH_3). The sample was dry-loaded onto silica (1 mL) and purified by automated FCC (5:94.9:0.1 to 30:69.9:0.1, MeOH:EtOAc:NH_3) using a 12 g cartridge. Fractions containing product **130** were combined and the solvent was removed *in vacuo*. The product (122 mg, 60% isolated yield) was analysed by NMR in MeOD. M.p. >90°C (decomposes). $[\alpha]_D^{25}$= +38.75 (c=1.2, MeOH). IR (cm^{-1}) 3400-3200 (O-H), 3190-3080 (N-H), 2950-2850 (C-H), 1672 (C=N).^{1}H NMR (600 MHz, CD_3OD) δ 7.64 – 7.55 (m, 4 H, Si<u>Ph</u>), 7.34 – 7.23 (m, 6 H, Si<u>Ph</u>), 5.75 (d, *J* = 5.4 Hz, 1 H, H-(C1′)); 4.91 (t, *J* = 5.4 Hz, 1 H, H-(C2′), *peak partially obscured by residual solvent peak*); 4.08 (dd, *J* = 9.2, 5.4 Hz, 1 H, H-(C3′)); 3.70 (ABX, *J* = 11.7, 1.9 Hz, 1 H, H^{a}-(C5′)); 3.82 (ABX, *J* = 11.7, 4.5 Hz, 1 H, H^{b}-(C5′)); 3.58 – 3.54 (m, 1 H, H-(C4′)); 0.92 (s, 9 H, *t*-Bu). ^{13}C NMR (151 MHz, CD_3OD) δ_C 166.2 (C2); 136.8, 136.7, 134.5, 134.4, 130.9, 128.9 (Si<u>Ph</u>); 95.3 (C1′); 84.7 (C2′); 80.0 (C4′); 72.1 (C3′); 63.6 (C5′); 27.3, 20.1 (*t*-Bu). MS (EI) (*m/z*): [M] calcd for formula $C_{22}H_{28}N_2O_4Si$, 412.18183; found, 412.180474.

*3′-O-Benzoyl 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl 1/2-N-benzoyl-aminooxazoline (**131**)*



131a: R_1=Bz, R_2=H
131b: R_1=H, R_2=Bz

A stirred solution of 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl aminooxazoline (**130**; 3.38 g, 8.19 mmol) in pyridine (75 mL) was kept under argon atmosphere and cooled to 0°C. Benzoic anhydride (5.56 g, 24.6 mmol) was added, followed by DMAP (catalytic amount). The resulting solution was stirred under argon at RT for 21 h. The solvent was removed *in vacuo* and the resulting residue dissolved in CH_2Cl_2 (30 mL), then washed with water (3×30 mL), brine (30 mL) and dried by saturating in $MgSO_4$. The solution was filtered and the solvent removed *in vacuo*. The sample was then purified by automated FCC (5:95 to 40:60, EtOAc:Hexane) using a 40 g cartridge. Fractions containing product 3′-*O*-benzoyl 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl 1/2-*N*-benzoyl-aminooxazoline (**131**) were combined and the solvent was removed *in vacuo*. The product **131** (1.86 g, 38%) was analysed by TLC and showed the presence of an alternative product **131**b. Another purification by automated FCC (5:95 to 40:60, EtOAc:Hexane) using a 40 g cartridge was set up in order to attempt to separate the two compounds. Fractions containing pure product **131**a were combined and the solvent was removed *in vacuo*. The product (370 mg, 7% over two purifications) was analysed by NMR in $CDCl_3$. Fractions containing a mixture of **131**a and **131**b were combined and the solvent was removed *in vacuo*. The mixture (1.27 g) was analysed by ^{1}H NMR in $CDCl_3$.

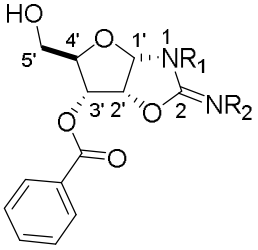
3′-*O*-Benzoyl 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl 1-*N*-benzoyl-aminooxazoline (**131**a):

M.p. >65°C (decomposes). $[\alpha]_D^{25}$ = +80.04 (c=1, CH_2Cl_2). IR (cm^{-1}) 3090 (N-H), 2980-2750 (C-H), 1728-1660 (C=O, C=N). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.93-7.89 (m, 2 H, SiPh+COPh); 7.82-7.79 (m, 2 H, SiPh+COPh); 7.69-7.66 (m, 2 H, SiPh+COPh); 7.65-7.61 (m, 4 H, SiPh+COPh); 7.57-7.52 (m, 2 H, SiPh+COPh); 7.46-7.25 (m, 8 H, SiPh+COPh); 6.57 (d, *J* = 5.3 Hz, 1 H, H-(C1′)); 5.57-5.54 (m, 1 H, H-(C3′), *peak obscured by H-(C2') peak*); 5.54-5.50 (m, 1 H, H-(C2′), *peak obscured by H-(C3') peak*); 4.50-4.45 (m, 1 H, H-(C4′)); 4.02 (ABX, *J* = 11.8, 2.6 Hz, 1 H, H^a-(C5′)); 3.89 (ABX, *J* = 11.8, 2.6 Hz, 1 H, H^b-(C5′)); 1.06 (s, 9 H, *t*-Bu). ^{13}C NMR (151 MHz, $CDCl_3$) δ_C 174.8 (COPh); 169.0, 165.6 (COPh); 149.4 (C2), 135.7, 130.0, 135.6, 133.7, 132.8, 132.6, 130.1, 129.7, 129.3, 128.6, 128.3, 128.1, 128.0, 128.0 (SiPh+COPh); 88.5 (C1′); 81.6 (C4′); 78.2 (C2′); 77.4 (C3′); 62.7 (C5′); 26.9, 19.3 (*t*-Bu). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{36}H_{37}N_2O_6Si^+$, 621.7845; found, 621.2452.

3′-*O*-benzoyl 5′-tert-butyldiphenylsilyl-ether-D-ribofuranosyl 2-*N*-benzoyl-aminooxazoline (**131**b) (*partial assignment*):

^{1}H NMR (600 MHz, $CDCl_3$) δ 6.88 (d, *J* = 9.8 Hz, 1 H, H-(N1)); 6.60 (dd, *J* = 9.8, 3.8 Hz, 1 H, H-(C1′)); 5.36-5.31 (m, 1 H); 5.22 (dd, *J* = 6.8, 3.8 Hz, 1 H); 4.36 (t, *J* = 2.5 Hz); 4.02 (ABX, *J* = 11.7, 2.3 Hz, 1 H, H^b-(C5′)); 3.88 (ABX, *J* = 11.7, 3.0 Hz, 1 H, H^b-(C5′)); 1.11 (s, 9 H, *t*-Bu).

*3′-O-Benzoyl-D-ribofuranosyl 1/2-N-benzoyl-aminooxazoline (**128**)*



128a: R_1=Bz, R_2=H
128b: R_1=H, R_2=Bz

To a stirred solution of 3′-*O*-benzoyl 5′-tert-butyldiphenylsilyl-ether D-ribofuranosyl 1/2-*N*-benzoyl-aminooxazoline (**131**a/**131**b; 1.26 g, 2.04 mmol) in CH_2Cl_2 (20 mL), cooled to 0°C, was slowly added TBAF (2.04 mL 1M in THF,

2.04 mmol). The resulting solution was left stirring overnight after which the solvent was removed *in vacuo* and purified by automated FCC (1:9 to 1:0 EtOAc:Hexane) on a 40 g cartridge. Fractions containing product 3′-*O*-benzoyl-D-ribofuranosyl 1-*N*-benzoyl-aminooxazoline (**128**a) were combined and the solvent was removed *in vacuo*. The product (206 mg, 27% yield) was analysed by NMR in $CDCl_3$. Fractions containing a mixture of **128**a and **128**b were combined and the solvent was removed *in vacuo*. The mixture (282 mg, 35% yield) was analysed by ^{1}H NMR analysis in $CDCl_3$.

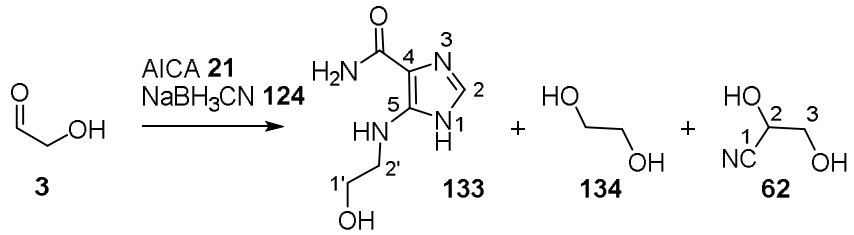
3′-*O*-Benzoyl-D-ribofuranosyl 1-*N*-benzoyl-aminooxazoline (**128**a):
M.p. >70°C (decomposes). $[\alpha]_D^{25}$= -165.71 (c=1, CH_2Cl_2). IR (cm^{-1}) 3450-3190 (O-H, N-H), 3090-2850 (C-H), 1724-1584 (C=O, C=N).^{1}H NMR (600 MHz, $CDCl_3$) δ 8.24-8.20 (m, 2 H, CO<u>Ph</u>); 8.13-8.08 (m, 2 H, CO<u>Ph</u>); 7.65-7.35 (m, 6 H, CO<u>Ph</u>); 6.09 (d, *J* = 5.3 Hz, 1 H, H-(C1′)); 5.45 (t, *J* = 5.3 Hz, 1 H, H-(C2′)); 5.30 (dd, *J* = 9.4, 5.3 Hz, 1 H, H-(C3′)); 4.20 (dt, *J* = 9.4, 2.5 Hz, 1 H, H-(C4′)); 4.07 (ABX, *J* = 13.1, 2.5 Hz, 1 H, H^a-(C5′)); 3.82 (ABX, *J* = 13.1, 2.5 Hz, 1 H, H^b-(C5′)). ^{13}C NMR (151 MHz, $CDCl_3$) δ_C 179.2 (CO<u>Ph</u>); 166.5, 166.0 (<u>C</u>OPh); 136.2, 134.0, 132.7, 130.3, 129.8, 128.7, 128.6, 128.3 (CO<u>Ph</u>); 87.7 (C1′); 78.4 (C2′); 76.9 (C4′); 71.2 (C3′); 59.8 (C5′). HRMS (*m/z*): $[M+H^+]^+$ calcd for formula $C_{20}H_{19}N_2O_6^+$, 383.1238; found, 383.1245.

3′-*O*-benzoyl-D-ribofuranosyl 2-*N*-benzoyl-aminooxazoline (**128**b) (*partial assignment*):
^{1}H NMR (600 MHz, $CDCl_3$) δ 6.96 (d, *J* = 9.8 Hz, 1 H, H-(N1)); 6.39 (dd, *J* = 9.8, 3.8 Hz, 1 H, H-(C1′)); 5.37 (d, *J* = 6.8 Hz, 1 H); 5.21 (dd, *J* = 6.8, 3.8 Hz, 1 H); 4.40 (t, *J* = 2.7 Hz, 1 H); 4.03 (ABX, *J* = 11.9, 2.7 Hz,1 H, H^a-(C5′)); 3.86 (ABX, *J* = 11.9, 2.7 Hz, 1 H, H^b-(C5′)).

9.7.2. Reductive aminations

General procedure for the reductive amination of aldehydes

To a solution of aldehyde (0.25 mmol) in water (2 mL) was added AICA **21** (1-6 eq). The solution was gently heated to dissolve solids, and the solution was adjusted to the desired pH (3-9). A solution of sodium cyanoborohydride (**124**; 47 mg, 0.75 mmol) was made in water (0.5 mL) at the same pH. The solutions were combined and the resulting solution adjusted to the desired pH, and monitored periodically. The reaction progress was monitored by ^{1}H NMR analysis.



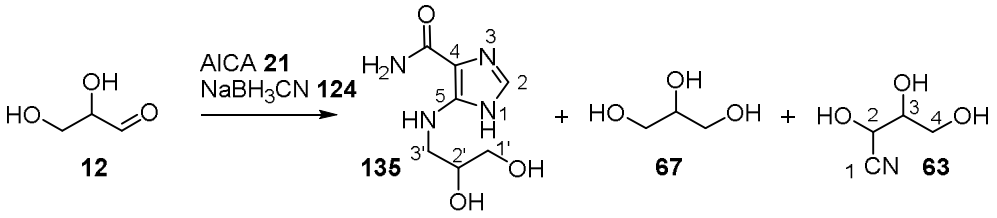
^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 6) (*partial assignment*)

5-((2-Hydroxyethyl)amino)-imidazole-4-carboxamide (**133**): 7.30 (s, 1 H, H-(C2)); 3.68 (t, *J* = 5.4 Hz, 2 H, H-(C2′)); 3.34 (t, *J* = 5.4 Hz, 2 H, H-(C1′)).

4-amino-imidazole-5-carboxamide (**21**): 7.26 (s, 1 H, H-(C2)).

Ethylene glycol (**134**): 3.61 (s, 4 H).

2,3-Dihydroxypropanenitrile (**62**): 4.65 (t, *J* = 4.8 Hz, 1 H, H-(C2)); 3.77 (d, *J* = 4.8 Hz, 2 H, H-(C3)).



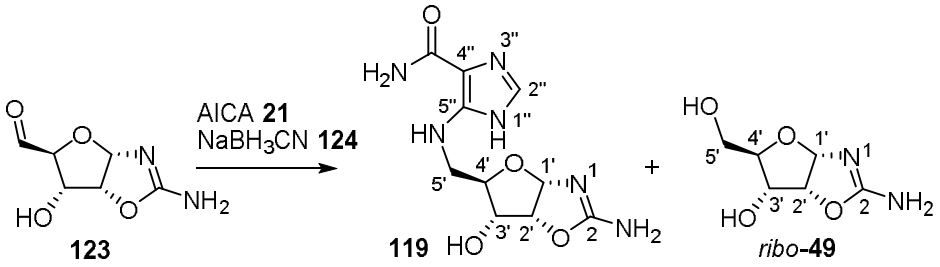
^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7)

5-((2,3-Dihydroxypropyl)amino)-imidazole-4-carboxamide (**135**): 7.18 (s, 1 H, H-(C2)); 3.73-3.57 (m, 1 H, H-(C2′)); 3.51-3.45 (ABX, *J* = 4.4, 1 H, H^a-(C1′), *peak obscured by **67***); 3.53 (ABX, *J* = 11.7, 6.8 Hz, 1 H, H^b-(C1′), *peak obscured by **67***); 3.21 (ABX, *J* = 13.9, 4.4 Hz, 1 H, H^a-(C3′)); 3.06 (ABX, *J* = 13.9, 6.8 Hz, 1 H, H^b-(C3′)). ^{13}C NMR (150 MHz, H_2O/D_2O 9:1) δ_C 167.7 ($H_2N\underline{C}O$-C4); 149.9 (C5); 133.6 (C2); 107.2 (C4); 71.3 (C2′); 63.9 (C1′); 46.8 (C3′).

4-amino-imidazole-5-carboxamide (**21**): 7.14 (s, 1 H, H-(C2)). ^{13}C NMR (150 MHz, H_2O/D_2O 9:1) δ_C 168.2 ($H_2N\underline{C}O$-C4); 146.7 (C5); 132.8 (C2); 109.5 (C4).

Glycerol (**67**): 3.64-3.59 (m, 1 H, H-(C2)); 3.37 (ABX, *J* = 11.8, 4.2, 2 H, H^a-(C1/C3)); 3.47 (ABX, *J* = 11.8, 6.6 Hz, 2 H, H^b-(C1/C3)). ^{13}C NMR (150 MHz, H_2O/D_2O 9:1) δ_C 72.8 (C2); 63.2 (C1/C3).

2,3,4-Trihydroxybutanenitrile (**63**): 4.54 (d, *J* = 5.7, 1 H, H-(C2)); 3.81-3.76 (m, 1 H, H-(C3)); 3.56-3.52 (m, 2 H, H-(C4)).



^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 5)

5′-(5-amino-imidazole-4-carboxamide)-D-ribofuranosyl aminooxazoline (**119**): 7.21 (s, 1 H, H-(C2″)); 5.86 (d, *J* = 5.2 Hz, 1 H, H-(C1′), *peak overlapping with H-(C1') ribo-**49***); 5.23 (t, *J* = 5.2 Hz, 1 H, H-(C2′), *peak overlapping with H-(C2') ribo-**49***); 4.07 (dd, *J* = 9.5, 5.3 Hz, 1 H, H-(C3′), *peak overlapping with H-(C3') ribo-**49***); 3.79-3.77 (H-(C4′), *peak obscured by H^a-(C5') ribo-**49***); 3.49 (ABX, *J* = 14.7, 2.8 Hz, 1 H, H^a-(C5′)); 3.33 (ABX, *J* = 14.7, 5.3 Hz, 1 H, H^b-(C5′)).

4-amino-imidazole-5-carboxamide (**21**): 7.23 (s, 1 H, H-(C2)). ^{13}C NMR (150 MHz, H_2O/D_2O 9:1) δ_C 168.2 ($H_2N\underline{C}O$-C4); 146.7 (C5); 132.8 (C2); 109.5 (C4).

D-Ribofuranosyl aminooxazoline (*ribo*-**49**): 5.86 (d, *J* = 5.3 Hz, 1 H, H-(C1′), *peak overlapping with H-(C1') **119***); 5.25 (t, *J* = 5.3 Hz, 1 H, H-(C2′), *peak overlapping with H-(C2') **119***); 4.10 (dd, *J* = 9.7, 5.3 Hz, 1 H, H-(C3′), *peak overlapping with H-(C3') **119***); 3.78-3.75 (H^a-(C5′), *peak obscured by H-(C4') **119***); 3.68 (ddd, *J* = 9.7, 4.7, 2.1 Hz, 1 H, H-(C4′)); 3.56 (dd, *J* = 12.8, 4.7 Hz, 1 H, H^b-(C5′)).

Reduction side-products ethylene glycol (**134**), glycerol (**67**) and riboaminooxazoline (*ribo*-**49**) were confirmed by spiking with authentic samples (*Figures 7.8*, *7.10* and *7.11* respectively).

Tables showing the results of the reductive amination studies of glycolaldehyde (**3**), glyceraldehyde (**12**) and 5'-oxo-riboaminooxazoline (**123**) with AICA **21** at RT and a range of pH. The percentage of each compound in solution is based on ^{1}H NMR integration relative to other species in solution.

3:**21**:**124**	solvent	pH	Time (h)	^{1}H NMR integration relative to other species integrated (%)			
				3	**133**	**134**	**62**
1:1:3	D_2O	4	1	7.4	42.1	31.7	0
1:1:3	H_2O	5	1	0	39.6	54.5	5.9
1:1:3	H_2O	6	1	2.4	24.5	47.4	25.7
1:3:3	H_2O	5	1	0	74.5	25.5	0

*Table 9.28: Reductive amination of **3** (100mM) and **21** (HCl salt; 1-3 eq) at RT.*

3:**21**:**124**	solvent	pH	Time (h)	^{1}H NMR integration relative to other species integrated (%)			
				12	**133**	**134**	**62**
1:3*:3	H_2O	3	1.5	0	69.2	30.8	0.0
1:3:3	Acetate buffer	4.7	1.5	0	62.1	34.2	3.7
1:3:3	Pi buffer	7	1.5	0	29.4	50.0	20.6
1:1*:3	H_2O	4.5	1	0	50.3	37.7	11.1
1:3:3	Acetate buffer	4.7	1.5	0	62.1	34.2	3.7
1:6:3	Acetate buffer	4.7	2.5	0	68.0	32.0	0.0

*Table 9.29: Reductive amination of **12** (100mM) and **21** (*HCl salt; 1-6 eq) at RT. Acetate buffer: 1M acetate buffer pH 4.7; P_i buffer: 1M P_i buffer pH 7.*

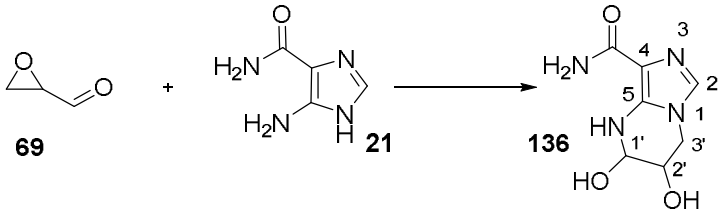
Time (h)	^{1}H NMR integration relative to other species integrated (%)			
	123	**119**	*ribo*-**49**	other*
0.75	0	36.2	51.1	12.7
1.25	0	37.7	49.4	12.8
2	0	36.9	51.3	11.8
3	0	39.7	41.3	19.0

*Table 9.30: Reductive amination of **123** (100mM) and **21** (6 eq) with **124** (6 eq) in 1M acetate buffer at pH 4.7, over time. *unidentified minor species. The NMR analysis of last time-point (3 h) was performed on a lyophilised aliquot dissolved in D_2O.*

General procedure for opening of glycidaldehyde (69**) with various nucleophiles**

A solution of nucleophile (0.42 mmol) in D_2O (1 mL) was prepared and the solution was adjusted to the desired pH (4-8). Glycidaldehyde (**69**; 10 mg, 0.14 mmol) was added and the pH adjusted if necessary, using 1M HCl/1M NaOH. The reaction progress was monitored by ^{1}H NMR analysis.

*AICA **21** opening*



The procedure described above (*General procedure for the opening of glycidaldehyde (**69**) with various nucleophiles*) was followed at pH 4.4 and 7.7. AICA **21** (3 eq) was gently heated to dissolve before addition of **69**.

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

Imidazole **136** (major isomer): 7.19 (s, 1 H, H-(C2)); 4.96 (d, *J* = 3.0 Hz, 1 H, H-(C1')); 4.24 (q, *J* = 3.0 Hz, 1 H, H-(C2')); 4.21 (ABX, *J* = 13.8, 3.0 Hz, 1 H, H^a-(C3')); 4.15 (ABX, *J* = 13.8, 3.0 Hz, 1 H, H^b-(C3')). ^{13}C NMR (151 MHz, D_2O) δ_C 168.8 (C=O); 140.1 (C5); 132.2 (C2); 109.8 (C4); 74.2 (C1′); 63.3 (C2′); 43.8 (C3′). MS (ES+) (*m/z*): $[M+H^+]^+$ calcd for formula $C_7H_{11}N_4O_3^+$, 199.0826; found, 199.0710.

Imidazole **136** (minor isomer): 4.99 (d, *J* = 2.5 Hz, 1 H, H-(C1')).

Tables showing the results of the reaction of glycidaldehyde (**69**) and AICA **21** in water at RT. The percentage of each compound in solution is based on ^{1}H NMR integration relative to other species in solution.

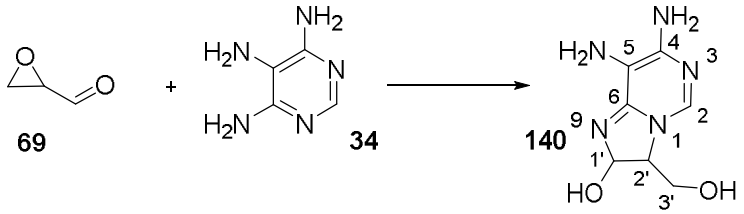
Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**136**	other*
0.2	73.0	3.6	23.4
1	53.5	19.8	26.7
3	25.3	44.4	30.3
6	14.3	47.2	38.5
24	0.0	56.2	43.8
66	0.0	55.6	44.4

*Table 9.31: Reaction of **69** (140mM) and **21** (3 eq) at pH 4.4. *unidentified species.*

Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**136**	other*
0.2	76.3	4.6	19.1
1	45.2	11.8	43.0
3	27.5	48.5	24.0
6	20.1	50.1	29.8
24	0.0	73.5	26.5
66	0.0	79.4	20.6

*Table 9.32: Reaction of **69** (140mM) and **21** (3 eq) at pH 7.7. *unidentified species.*

*Triaminopyrimidine **34** opening*



The procedure described above (*General procedure for the opening of glycidaldehyde (**69**) with various nucleophiles*) was followed at pH 7 (140mM) with 4,5,6-triaminopyrimidine (**34**; 3 eq). The reaction was also carried out more dilute (35mM; 3 eq **34**) at a range of pH (3-7).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7)

Triaminopyrimidine **140** (major isomer): 8.14 (s, 1 H, H-(C2)); 5.51 (d, *J* = 1.7 Hz, 1 H, H-(C1')); 4.61 (ddd, *J* = 6.1, 4.1, 1.7 Hz, 1 H, H-(C2')); 3.95 (ABX, *J* = 12.6, 4.1 Hz, 1H, H^a-(C3')); 3.87 (ABX, *J* = 12.6, 6.1 Hz, 1 H, H^b-(C3')). ^{13}C NMR (151 MHz, D_2O) δ_C 156.2 (C4); 146.8 (C6); 142.2 (C2); 103.1 (C5); 82.0 (C1′); 70.2 (C2′); 61.2 (C3′).

Triaminopyrimidine **140** (minor isomer): 8.21 (s, 1 H, H-(C2)); 5.78 (d, *J* = 6.6 Hz, 1 H, H-(C1')); 4.69 (ddd, *J* = 6.6, 4.2, 2.2 Hz, 1 H, H-(C2')); 4.15 (ABX, *J* = 12.7, 4.2 Hz, 1H, H^a-(C3')); 4.05 (ABX, *J* = 12.7, 6.6 Hz, 1 H, H^b-(C3')). ^{13}C NMR (151 MHz, D_2O) δ_C 156.1 (C4); 146.8 (C6); 142.1 (C2); 102.7 (C5); 80.7 (C1′); 65.5 (C2′); 58.5 (C3′).

Tables showing the results of the reaction of glycidaldehyde (**69**) and 4,5,6-triaminopyrimidine (**34**) in water at RT. The percentage of each compound in solution is based on ^{1}H NMR integration relative to other species in solution.

Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**140**	other*
0.25	28.7	30.6	40.7
1	4	62.4	37.6
2	0	40.6	59.4

*Table 9.33: Reaction of **69** (140mM) and **34** (3 eq) at pH 6. *unidentified species.*

Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**140**	other*
2	95	3	2
3	82	4	14

*Table 9.34: Reaction of **69** (35mM) and **34** (3 eq) at pH 3. *unidentified species.*

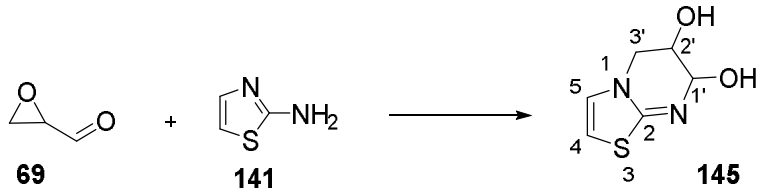
Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**140**	other*
2	10.3	46.6	43.1
3	0.7	39	60.3

*Table 9.35: Reaction of **69** (35mM) and **34** (3 eq) at pH 5. *unidentified species.*

Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**140**	other*
2	0	34	66
3	0	32	68

*Table 9.36: Reaction of **69** (35mM) and **34** (3 eq) at pH 7. *unidentified species.*

*2-Aminothiazole (**141**) opening*



69 **141** **145**

The procedure described above (*General procedure for the opening of glycidaldehyde (**69**) with various nucleophiles*) was followed at pH 5 and 7 with 2-aminothiazole (**141**; 1 eq).

^{1}H NMR (600 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

Thiazole **145** (major isomer): 7.21 (d, *J* = 4.3 Hz, 1 H, H-(C5)); 6.94 (d, *J* = 4.3 Hz, 1 H, H-(C4)), 5.04 (dd, *J* = 2.6, 1.3 Hz, 1 H, H-(C1′)); 4.28-4.25 (H-(C2'), *peak overlaps with H^a-(C3')*); 4.25-4.22 (H^a-(C3'), *peak overlaps with H-(C2')*); 4.25 (ABX, *J* = 14.5, 2.6 Hz, 1 H, H^b-(C3')). ^{13}C NMR (151 MHz, D_2O) δ_C 164.2 (C2); 130.5 (C5); 108.7 (C4); 74.7 (C1′); 62.1 (C2′); 48.3 (C3′).

Thiazole **145** (minor isomer) 5.18 (br. s., 1 H, H-(C1′)).

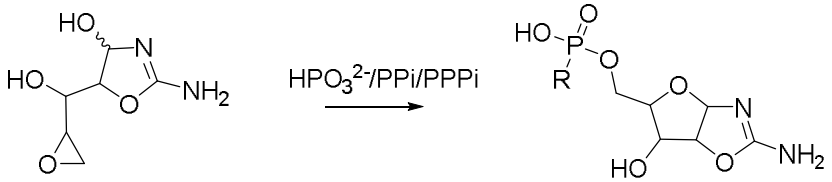
Tables showing the results of the reaction of glycidaldehyde (**69**) and 2-aminothiazole (**141**) in water at RT. The percentage of each compound in solution is based on ^{1}H NMR integration relative to other species in solution.

Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**145**	other*
0.3	83.7	6.6	9.3
2.5	60.5	28.1	11.4
6.5	44.9	50.4	4.7
11	39.8	55.7	4.5

*Table 9.37: Reaction of **69** (140mM) and **141** (1 eq) at pH 5. *unidentified species.*

Time (h)	^{1}H NMR integration relative to other species integrated (%)		
	69	**145**	other*
0.3	71.8	9.3	19
2.5	43.9	49.2	6.9
6.5	24.9	71.4	3.7
11	18.3	81.1	0.6

*Table 9.38: Reaction of **69** (140mM) and **141** (1 eq) at pH 7. *unidentified species.*



R= H **149**
PO_4^{2-} **147**
$HP_2O_7^{2-}$ **148**

Pyrophosphate opening

Sodium pyrophosphate tetrabasic (187.3 mg, 0.42 mmol) was added to a 1 mL aliquot of the reaction described previously (*Preparative synthesis of oxazolines **97**, p. 312*). The solution was adjusted to pH 7 and monitored over 11 days. The reaction was monitored by ^{1}H and ^{31}P NMR over time.

^{31}P (121 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

5′-diphosphate aminooxazolines (**147**): -7.08, -7.36, -7.53, -7.83[lx].

Triphosphate opening

Sodium triphosphate pentabasic (154.5 mg, 0.42 mmol) was added to a 1 mL aliquot of the reaction described previously (*Preparative synthesis of oxazolines*

[lx] It was not possible to assign which of the four diastereomers each phosphorus signal corresponded to, so they are listed here unassigned.

97, p.312). The solution was adjusted to pH 7 and monitored over 20 days. The reaction was monitored by ^{1}H and ^{31}P NMR over time.

^{31}P (162 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

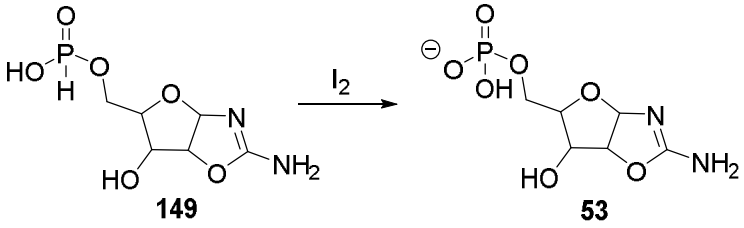
5′-triphosphate aminooxazolines (**148**): -10.05, -10.17, -10.28, -10.43[lx].

Phosphite opening

Sodium phosphite dibasic (90.7 mg, 0.42 mmol) was added to a 1 mL aliquot of the reaction described previously (*Preparative synthesis of oxazolines **97**, p.312*). The solution was adjusted to pH 7 and monitored over 11 days.

^{31}P (121 MHz, H_2O/D_2O 9:1) δ ppm (shifts are pH dependent, given here for pH 7) (*partial assignment*)

5′-phosphite aminooxazolines (**149**): 9.81, 9.57, 9.53, 9.31[lx].



149 **53**

Oxidation of crude 5'-phosphite aminooxazolines (**149**) was carried out on an aliquot (0.5 mL) of the above reaction (*Phosphite opening*) using iodine (53.3 mg, 0.21 mmol). The reaction was monitored by ^{1}H and ^{31}P NMR over time. After 20 days the reaction was deemed to have stabilised and another equivalent of iodine was added and the reaction was monitored further for 13 days.

^{31}P (162 MHz, H_2O/D_2O 9:1) δ ppm (*partial assignment*)

5′-phosphate aminooxazolines (**53**): -0.02, -0.06, -0.12, -0.21[lx].

10. References

1. Margulis, L. & Sagan, D. *What is Life?* (University of California Press, 2000).
2. Darwin, C. *The Origin Of Species*. (1872).
3. Darwin, C. & Wallace, A. On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *J. Proc. Linn. Soc. Lond. Zool.* 3, 45–62 (1858).
4. Green, J. H. S. William Charles Wells, F.R.S. (1757–1817). *Nature* 179, 997–999 (1957).
5. Peretó, J., Bada, J. L. & Lazcano, A. Charles Darwin and the Origin of Life. *Orig. Life Evol. Biospheres* 39, 395–406 (2009).
6. Schrödinger, E. *What is Life?: With Mind and Matter and Autobiographical Sketches*. (Cambridge University Press, 1992).
7. Franklin, R. E. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* 171, 740–741 (1953).
8. Franklin, R. E. & Gosling, R. G. Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* 172, 156–157 (1953).
9. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738 (1953).
10. Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* 171, 738–740 (1953).
11. Deamer, D. W. & Fleischaker, G. R. *Origins of life: the central concepts*. (Jones and Bartlett Publishers, 1994).
12. Gamow, G. Possible Relation between Deoxyribonucleic Acid and Protein Structures. *Nature* 173, 318–318 (1954).
13. Crick, F. H. C., Leslie Barnett, F. R. S., Brenner, S. & Watts-Tobin, R. J. General Nature of the Genetic Code for Proteins. *Macmillan Journals Limited* 129, 1227–1232 (1961).
14. Nirenberg, M. W. & Matthaei, J. H. The Dependence Of Cell-Free Protein Synthesis in E. Coli Upon Naturally Occurring or Synthetic Polyribonucleotides. *Proc. Natl. Acad. Sci.* 47, 1588–1602 (1961).
15. Nirenberg, M. & Leder, P. RNA Codewords and Protein Synthesis. *Science* 145, 1399–1407 (1964).
16. Eiserling, F. *et al.* Polyribosomes and DNA-dependent Amino Acid Incorporation in Escherichia Coli Extracts. *J. Mol. Biol.* 10, 536–540 (1964).
17. Nirenberg, M. *et al.* RNA Codewords and Protein Synthesis, VII. On the General Nature of the RNA Code. *Proc. Natl. Acad. Sci. U. S. A.* 53, 1161–1168 (1965).

18. Bernfield, M. R. & Nirenberg, M. W. RNA Codewords and Protein Synthesis. *Science* 147, 479–484 (1965).
19. Kellogg, D. A., Doctor, B. P., Loebel, J. E. & Nirenberg, M. W. RNA Codons and Protein Synthesis, IX. Synonym Codon Recognition by Multiple Species of Valine-, Alanine-, and Methionine-sRNA. *Proc. Natl. Acad. Sci. U. S. A.* 55, 912–919 (1966).
20. Khorana, H. G. Synthesis in the Study of Nucleic Acids. The Fourth Jubilee Lecture. *Biochem. J.* 109, 709–725 (1968).
21. Bladen, H. A., Byrne, R., Levin, J. G. & Nirenberg, M. W. An electron microscopic study of a DNA-ribosome complex formed in vitro. *J. Mol. Biol.* 11, 78–83 (1965).
22. Lodish, H. *et al*. *Molecular Cell Biology*. (2000).
23. Crick, F. H. C. The Origin of The Genetic Code. *J. Mol. Biol.* 38, 367–379 (1968).
24. Borsenberger, V. *et al*. Exploratory Studies to Investigate a Linked Prebiotic Origin of RNA and Coded Peptides. *Chem. Biodivers.* 1, 203–246 (2004).
25. Orgel, L. E. Evolution of the Genetic Apparatus. *J. Mol. Biol.* 38, 381–393 (1968).
26. Woese, C. R. *The Genetic Code: the Molecular Basis for Genetic Expression*. (Harper & Row, 1967).
27. Wong, J. T.-F. A Co-Evolution Theory of the Genetic Code. *Proc. Natl. Acad. Sci. U. S. A.* 72, 1909–1912 (1975).
28. Orgel, L. E. The Origin of Polynucleotide-Directed Protein Synthesis. *J. Mol. Evol.* 29, 465–474 (1989).
29. Wong, J. T.-F. Origin of Genetically Encoded Protein Synthesis: a Model Based on Selection for RNA Peptidation. *Orig. Life Evol. Biosph.* 21, 165–176
30. Feng, D.-F., Cho, G. & Doolittle, R. F. Determining Divergence Times With a Protein Clock: Update And Reevaluation. *Proc. Natl. Acad. Sci.* 94, 13028–13033 (1997).
31. Orgel, L. E. The Origin of Life—a Review of Facts and Speculations. *Trends Biochem. Sci.* 23, 491–495 (1998).
32. Eschenmoser, A. Etiology of Potentially Primordial Biomolecular Structures: from Vitamin B12 to the Nucleic Acids and an Inquiry into the Chemistry of Life's Origin: a Retrospective. *Angew. Chem. Int. Ed. Engl.* 50, 12412–12472 (2011).
33. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature* 459, 239–242 (2009).
34. Joyce, G. F. The antiquity of RNA-based Evolution. *Nature* 418, 214–221 (2002).

35. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing Life. *Nature* 409, 387–390 (2001).
36. Miller, S. L. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science* 117, 528–529 (1953).
37. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common Origins of RNA, Protein and Lipid Precursors in a Cyanosulfidic Protometabolism. *Nat. Chem.*, (2015).
38. Sutherland, J. D. The Origin of Life—Out of the Blue. *Angew. Chem. Int. Ed.* 55, 104-121 (2016).
39. Buick, R. Carbonaceous filaments from North Pole, Western Australia: Are they Fossil Bacteria in Archaean Stromatolites? *Precambrian Res.* 24, 157–172 (1984).
40. Moorbath, S. Palaeobiology: Dating Earliest Life. *Nature* 434, 155–155 (2005).
41. Tera, F., Papanastassiou, D. A. & Wasserburg, G. J. Isotopic Evidence for a Terminal Lunar Cataclysm. *Earth Planet. Sci. Lett.* 22, 1–21 (1974).
42. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the Cataclysmic Late Heavy Bombardment Period Of The Terrestrial Planets. *Nature* 435, 466–469 (2005).
43. Hazen, R. M. Genesis: Rocks, Minerals, and the Geochemical Origin of Life. *Elements* 1, 135–137 (2005).
44. Oró, J. Comets and the Formation of Biochemical Compounds on the Primitive Earth. *Nature* 190, 389–390 (1961).
45. Thaddeus, P. The Prebiotic Molecules Observed in the Interstellar Gas. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 1681–1687 (2006).
46. Callahan, M. P. *et al.* Carbonaceous Meteorites Contain a Wide Range of Extraterrestrial Nucleobases. *Proc. Natl. Acad. Sci.* 108, 13995–13998 (2011).
47. Cronin, J. R. & Pizzarello, S. Enantiomeric Excesses in Meteoritic Amino Acids. *Science* 275, 951–955 (1997).
48. Cooper, G. *et al.* Carbonaceous Meteorites as a Source of Sugar-related Organic Compounds for the Early Earth. *Nature* 414, 879–883 (2001).
49. Wolman, Y., Haverland, W. J. & Miller, S. L. Nonprotein Amino Acids from Spark Discharges and Their Comparison with the Murchison Meteorite Amino Acids. *Proc. Natl. Acad. Sci.* 69, 809–811 (1972).
50. Miller, S. L. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science* 117, 528–529 (1953).
51. Trump, J. E. V. & Miller, S. L. Prebiotic Synthesis of Methionine. *Science* 178, 859–860 (1972).
52. Sagan, C. & Chyba, C. The Early Faint Sun Paradox: Organic Shielding of Ultraviolet-Labile Greenhouse Gases. *Science* 276, 1217–1221 (1997).

53. Sanchez, R., Ferris, J. & Orgel, L. E. Conditions for Purine Synthesis: Did Prebiotic Synthesis Occur at Low Temperatures? *Science* 153, 72–73 (1966).
54. Brack, A. *The Molecular Origins of Life: Assembling Pieces of the Puzzle*. (Cambridge University Press, 1998).
55. Clampin, M. The James Webb Space Telescope and its Capabilities for Exoplanet Science. *Proc. Int. Astron. Union* 6, 335–342 (2010).
56. Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R. & Zolotukhin, I. Defining and Cataloging Exoplanets: the Exoplanet.Eu Database. *Astron. Astrophys.* 532, A79 (2011).
57. The Extrasolar Planets Encyclopaedia. Available at: http://exoplanet.eu/. (Accessed: 22nd September 2016)
58. Schopf, J. W. Fossil evidence of Archaean life. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 869–885 (2006).
59. Oparin, A. I. *The Origin of Life*. (Courier Dover Publications, 2003).
60. Kasting, J. F. Earth's Early Atmosphere. *Science* 259, 920–926 (1993).
61. Crowe, S. A. *et al.* Atmospheric Oxygenation Three Billion Years Ago. *Nature* 501, 535–538 (2013).
62. Mojzsis, S. J., Harrison, T. M. & Pidgeon, R. T. Oxygen-isotope Evidence from Ancient Zircons for Liquid Water at the Earth's Surface 4,300 Myr ago. *Nature* 409, 178–181 (2001).
63. Orgel, L. E. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* 39, 99–123 (2004).
64. Monnard, P.-A., Apel, C. L., Kanavarioti, A. & Deamer, D. W. Influence of Ionic Inorganic Solutes on Self-Assembly and Polymerization Processes Related to Early Forms of Life: Implications for a Prebiotic Aqueous Medium. *Astrobiology* 2, 139–152 (2002).
65. Usher, D. A. Early Chemical Evolution of Nucleic Acids: a Theoretical Model. *Science* 196, 311–313 (1977).
66. Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S Ribosomal DNA Amplification for Phylogenetic Study. *J. Bacteriol.* 173, 697–703 (1991).
67. Woese, C. R. Interpreting the Universal Phylogenetic Tree. *Proc. Natl. Acad. Sci.* 97, 8392–8396 (2000).
68. Gaucher, E. A., Thomson, J. M., Burgan, M. F. & Benner, S. A. Inferring the Palaeoenvironment of Ancient Bacteria on the Basis of Resurrected Proteins. *Nature* 425, 285–288 (2003).
69. Nisbet, E. G. & Sleep, N. H. The Habitat and Nature of Early Life. *Nature* 409, 1083–1091 (2001).
70. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An Archaeal Origin of Eukaryotes Supports Only Two Primary Domains of Life. *Nature* 504, 231–236 (2013).

71. Glass, J. I. *et al.* Essential Genes of a Minimal Bacterium. *Proc. Natl. Acad. Sci. U. S. A.* 103, 425–430 (2006).
72. Hutchison, C. A. *et al.* Design and Synthesis of a Minimal Bacterial Genome. *Science* 351, aad6253 (2016).
73. Keller, M. A., Turchyn, A. V. & Ralser, M. Non□enzymatic Glycolysis and Pentose Phosphate Pathway□Like Reactions in a Plausible Archean Ocean. *Mol. Syst. Biol.* 10, 725 (2014).
74. Ralser, M. The RNA World and the Origin of Metabolic Enzymes. *Biochem. Soc. Trans.* 42, 985–988 (2014).
75. Wächtershäuser, G. Evolution of the First Metabolic Cycles. *Proc. Natl. Acad. Sci.* 87, 200–204 (1990).
76. Ledin, M. & Pedersen, K. The Environmental Impact of Mine Wastes — Roles of Microorganisms and their Significance in Treatment of Mine Wastes. *Earth-Sci. Rev.* 41, 67–108 (1996).
77. Cody, G. D. Geochemical Connections to Primitive Metabolism. *Elements* 1, 139–143 (2005).
78. Blöchl, E., Keller, M., Wachtershäuser, G. & Stetter, K. O. Reactions Depending on Iron Sulfide and Linking Geochemistry with Biochemistry. *Proc. Natl. Acad. Sci.* 89, 8117–8120 (1992).
79. Dörr, M. *et al.* A Possible Prebiotic Formation of Ammonia from Dinitrogen on Iron Sulfide Surfaces. *Angew. Chem. Int. Ed.* 42, 1540–1543 (2003).
80. Soai, K., Shibata, T., Morioka, H. & Choji, K. Asymmetric autocatalysis and amplification of enantiomeric Excess of a Chiral Molecule. *Nature* 378, 767–768 (1995).
81. Hafenbradl, D., Keller, M., Wächtershäuser, G. & Stetter, K. O. Primordial Amino Acids by Reductive Amination of α-oxo Acids in Conjunction with the Oxidative Formation of Pyrite. *Tetrahedron Lett.* 36, 5179–5182 (1995).
82. Huber, C. & Wächtershäuser, G. Primordial Reductive Amination Revisited. *Tetrahedron Lett.* 44, 1695–1697 (2003).
83. Russell, M. J., Nitschke, W. & Branscomb, E. The Inevitable Journey to Being. *Philos. Trans. R. Soc. B Biol. Sci.* 368, 20120254 (2013).
84. Lane, N., Allen, J. F. & Martin, W. How Did LUCA Make a Living? Chemiosmosis in the Origin of Life. *BioEssays* 32, 271–280 (2010).
85. Stetter, K. O. in *Ciba Foundation Symposium 202 - Evolution of Hydrothermal Ecosystems on Earth (And Mars?)* (eds. Organizer, G. R. B. & Goode, J. A.) 1–23 (John Wiley & Sons, Ltd., 2007).
86. Ludwig, K. A., Kelley, D. S., Butterfield, D. A., Nelson, B. K. & Früh-Green, G. Formation and Evolution of Carbonate Chimneys at the Lost City Hydrothermal Field. *Geochim. Cosmochim. Acta* 70, 3625–3645 (2006).
87. Kelley, D. S. *et al.* An Off-Axis Hydrothermal Vent Field Near the Mid-Atlantic Ridge at 30° N. *Nature* 412, 145–149 (2001).

88. Russell, M. J. & Hall, A. J. The Emergence of Life from Iron Monosulphide Bubbles at a Submarine Hydrothermal Redox and pH Front. *J. Geol. Soc.* 154, 377–402 (1997).
89. Sojo, V., Herschy, B., Whicher, A., Camprubí, E. & Lane, N. The Origin of Life in Alkaline Hydrothermal Vents. *Astrobiology* 16, 181–197 (2016).
90. Mulkidjanian, A. Y., Bychkov, A. Y., Dibrova, D. V., Galperin, M. Y. & Koonin, E. V. Origin of First Cells at Terrestrial, Anoxic Geothermal Fields. *Proc. Natl. Acad. Sci.* 109, E821–E830 (2012).
91. Baaske, P. *et al.* Extreme Accumulation of Nucleotides in Simulated Hydrothermal Pore Systems. *Proc. Natl. Acad. Sci.* 104, 9346–9351 (2007).
92. Budin, I., Bruckner, R. J. & Szostak, J. W. Formation of Protocell-like Vesicles in a Thermal Diffusion Column. *J. Am. Chem. Soc.* 131, 9628–9629 (2009).
93. Lazcano, A. An Answer in Search of a Question. *Astrobiology* 4, 469–471 (2004).
94. Jackson, J. B. Natural pH Gradients in Hydrothermal Alkali Vents Were Unlikely to Have Played a Role in the Origin of Life. *J. Mol. Evol.* 83, 1–11 (2016).
95. Higgs, P. G., & Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nature Rev. Gen.*, 16(1), 7-17 (2016).
96. Gilbert, W. Origin of Life: the RNA World. *Nature* 319, (1986).
97. Bansho, Y. *et al.* Importance of Parasite RNA Species Repression for Prolonged Translation-Coupled RNA Self-Replication. *Chem. Biol.* 19, 478–487 (2012).
98. Natsume, T. & Yoshimoto, M. Mechanosensitive Liposomes as Artificial Chaperones for Shear-Driven Acceleration of Enzyme-Catalyzed Reaction. *ACS Appl. Mater. Interfaces* 6, 3671–3679 (2014).
99. McCollom, T. M., Ritter, G. & Simoneit, B. R. T. Lipid Synthesis Under Hydrothermal Conditions by Fischer- Tropsch-Type Reactions. *Orig. Life Evol. Biosph.* 29, 153–166
100. Apel, C. L., Deamer, D. W. & Mautner, M. N. Self-Assembled Vesicles of Monocarboxylic Acids and Alcohols: Conditions for Stability and for the Encapsulation of Biopolymers. *Biochim. Biophys. Acta BBA - Biomembr.* 1559, 1–9 (2002).
101. Mansy, S. S. *et al.* Template-Directed Synthesis of a Genetic Polymer in a Model Protocell. *Nature* 454, 122–125 (2008).
102. Deamer, D., Dworkin, J. P., Sandford, S. A., Bernstein, M. P. & Allamandola, L. J. The First Cell Membranes. *Astrobiology* 2, 371–381 (2002).
103. Dworkin, J. P., Deamer, D. W., Sandford, S. A. & Allamandola, L. J. Self-Assembling Amphiphilic Molecules: Synthesis in Simulated Interstellar/Precometary Ices. *Proc. Natl. Acad. Sci.* 98, 815–819 (2001).

104. Hanczyc, M. M., Fujikawa, S. M. & Szostak, J. W. Experimental Models of Primitive Cellular Compartments: Encapsulation, Growth, and Division. *Science* 302, 618–622 (2003).
105. Strulson, C. A., Molden, R. C., Keating, C. D. & Bevilacqua, P. C. RNA Catalysis Through Compartmentalization. *Nat. Chem.* 4, 941–946 (2012).
106. Bowman, J. C., Lenz, T. K., Hud, N. V. & Williams, L. D. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. *Curr. Opin. Struct. Biol.* 22, 262–272 (2012).
107. Adamala, K. & Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis Inside Model Protocells. *Science* 342, 1098–1100 (2013).
108. Segré, D., Ben-Eli, D., Deamer, D. W. & Lancet, D. The Lipid World. *Orig. Life Evol. Biosph.* 31, 119–145
109. Sacerdote, M. G. & Szostak, J. W. Semipermeable Lipid Bilayers Exhibit Diastereoselectivity Favoring Ribose. *Proc. Natl. Acad. Sci.* 102, 6004–6008 (2005).
110. Lombard, J., López-García, P. & Moreira, D. The Early Evolution of Lipid Membranes and the Three Domains of Life. *Nat. Rev. Microbiol.* 10, 507–515 (2012).
111. Brack, A. & Orgel, L. E. β Structures of Alternating Polypeptides and their possible prebiotic significance. *Nature* 256, 383–387 (1975).
112. Ourisson, G. & Nakatani, Y. The terpenoid theory of the origin of Cellular Life: the Evolution of Terpenoids to Cholesterol. *Chem. Biol.* 1, 11–23 (1994).
113. Lee, D. H., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. Emergence of Symbiosis in Peptide Self-Replication Through a Hypercyclic Network. *Nature* 390, 591–594 (1997).
114. Li, Y. *et al.* Dipeptide Seryl-Histidine and Related Oligopeptides Cleave DNA, Protein, and a Carboxyl Ester. *Bioorg. Med. Chem.* 8, 2675–2680 (2000).
115. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A Self-Replicating Peptide. *Nature* 382, 525–528 (1996).
116. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science* 292, 1863–1876 (2001).
117. Miller, S. L. Which Organic Compounds Could Have Occurred on the Prebiotic Earth? *Cold Spring Harb. Symp. Quant. Biol.* 52, 17–27 (1987).
118. Glavin, D. P., Bada, J. L., Brinton, K. L. F. & McDonald, G. D. Amino acids in the Martian meteorite Nakhla. *Proc. Natl. Acad. Sci.* 96, 8835–8838 (1999).
119. Zhang, G. *et al.* Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution. *Cell* 98, 811–824 (1999).

120. Szostak, J. W. The Eightfold Path to Non-Enzymatic RNA Replication. *J. Syst. Chem.* 3, 1–14 (2012).
121. Steitz, T. A. & Steitz, J. A. A General Two-Metal-Ion Mechanism for Catalytic RNA. *Proc. Natl. Acad. Sci.* 90, 6498–6502 (1993).
122. Tamura, K. & Schimmel, P. Oligonucleotide-Directed Peptide Synthesis in a Ribosome- and Ribozyme-Free System. *Proc. Natl. Acad. Sci.* 98, 1393–1397 (2001).
123. Tamura, K. Origin of Amino Acid Homochirality: Relationship with the RNA World and Origin of tRNA Aminoacylation. *Biosystems* 92, 91–98 (2008).
124. Lowe, C. U., Rees, M. W. & Markham, R. Synthesis of Complex Organic Compounds from Simple Precursors: Formation of Amino-Acids, Amino-Acid Polymers, Fatty Acids and Purines from Ammonium Cyanide. *Nature* 199, 219–222 (1963).
125. Joyce, G. F. *et al.* Chiral Selection in poly(C)-Directed Synthesis of Oligo(G). *Nature* 310, 602–604 (1984).
126. Goldanskii, V. I., Avetisov, V. A. & Kuz'min, V. V. Chiral Purity of Nucleotides as a Necessary Condition of Complementarity. *FEBS Lett.* 207, 181–183 (1986).
127. Glavin, D. P. & Dworkin, J. P. Enrichment of the Amino Acid L-Isovaline by Aqueous Alteration on CI and CM Meteorite Parent Bodies. *Proc. Natl. Acad. Sci.* 106, 5487–5492 (2009).
128. Pizzarello, S., Zolensky, M. & Turk, K. A. Nonracemic Isovaline in the Murchison Meteorite: Chiral Distribution and Mineral Association. *Geochim. Cosmochim. Acta* 67, 1589–1595 (2003).
129. Viedma, C. Chiral Symmetry Breaking During Crystallization: Complete Chiral Purity Induced by Nonlinear Autocatalysis and Recycling. *Phys. Rev. Lett.* 94, 65504 (2005).
130. Kondepudi, D. K., Kaufman, R. J. & Singh, N. Chiral Symmetry Breaking in Sodium Chlorate Crystallizaton. *Science* 250, 975–976 (1990).
131. Viedma, C., Ortiz, J. E., Torres, T. de, Izumi, T. & Blackmond, D. G. Evolution of Solid Phase Homochirality for a Proteinogenic Amino Acid. *J. Am. Chem. Soc.* 130, 15274–15275 (2008).
132. Breslow, R. & Levine, M. S. Amplification of Enantiomeric Concentrations Under Credible Prebiotic Conditions. *Proc. Natl. Acad. Sci.* 103, 12979–12980 (2006).
133. Klussmann, M. *et al.* Thermodynamic Control of Asymmetric Amplification in Amino Acid Catalysis. *Nature* 441, 621–623 (2006).
134. Córdova, A., Engqvist, M., Ibrahem, I., Casas, J. & Sundén, H. Plausible Origins of Homochirality in the Amino Acid Catalyzed Neogenesis of Carbohydrates. *Chem. Commun.* 2047 (2005). doi:10.1039/b500589b

135. Ludlow, R. F. & Otto, S. Systems Chemistry. *Chem. Soc. Rev.* 37, 101 (2008).
136. Nitschke, J. R. Systems Chemistry: Molecular Networks Come of Age. *Nature* 462, 736–738 (2009).
137. Moore, P. B. & Steitz, T. A. The Involvement of RNA in Ribosome Function. *Nature* 418, 229–235 (2002).
138. Joyce, G. F. RNA Evolution and the Origins of Life. *Nature* 338, (1989).
139. Greider, C. W. & Blackburn, E. H. The Telomere Terminal Transferase of Tetrahymena is a Ribonucleoprotein Enzyme with Two Kinds of Primer Specificity. *Cell* 51, 887–898 (1987).
140. Lazcano, A., Guerrero, R., Margulis, L. & Oró, J. The Evolutionary Transition from RNA to DNA in Early Cells. *J. Mol. Evol.* 27, 283–290 (1988).
141. Burton, A. S. & Lehman, N. DNA Before Proteins? Recent Discoveries in Nucleic Acid Catalysis Strengthen the Case. *Astrobiology* 9, 125–130 (2009).
142. Neveu, M., Kim, H.-J. & Benner, S. A. The 'Strong' RNA World Hypothesis: Fifty Years Old. *Astrobiology* 13, 391–403 (2013).
143. White, H. B. Coenzymes as Fossils of an Earlier Metabolic State. *J. Mol. Evol.* 7, 101–104 (1976).
144. Visser, C. M. & Kellogg, R. M. Biotin. Its Place In Evolution. *J. Mol. Evol.* 11, 171–187
145. Benner, S. A., Ellington, A. D. & Tauer, A. Modern Metabolism as a Palimpsest of the RNA World. *Proc. Natl. Acad. Sci.* 86, 7054–7058 (1989).
146. Inoue, T. & Orgel, L. E. A Nonenzymatic RNA Polymerase Model. *Science* 219, 859–862 (1983).
147. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence Of Tetrahymena. *Cell* 31, 147–157 (1982).
148. Peebles, C. L. *et al.* A self-Splicing RNA Excises an Intron Lariat. *Cell* 44, 213–223 (1986).
149. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA Moiety of Ribonuclease P is the Catalytic Subunit of the Enzyme. *Cell* 35, 849–857 (1983).
150. Bartel, D. P., & Szostak, J. W. Isolation of New Ribozymes from a Large Pool of Random Sequences. *Tetrahedron Lett*, *32*, 4904 (1991).
151. Ekland, E. H., Szostak, J. W. & Bartel, D. P. Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences. *Science* 269, 364 (1995).
152. Lorsch, J. R. & Szostak, J. W. In Vitro Evolution of New Ribozymes with Polynucleotide Kinase Activity. *Nature* 371, 31–36 (1994).

153. Chen, X., Li, N. & Ellington, A. D. Ribozyme Catalysis of Metabolism in the RNA World. *Chem. Biodivers.* 4, 633–655 (2007).
154. Steitz, T. A. & Moore, P. B. RNA, the First Macromolecular Catalyst: the Ribosome is a Ribozyme. *Trends Biochem. Sci.* 28, 411–418 (2003).
155. Polacek, N., Gaynor, M., Yassin, A. & Mankin, A. S. Ribosomal Peptidyl Transferase Can Withstand Mutations At The Putative Catalytic Nucleotide. *Nature* 411, 498–501 (2001).
156. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* 289, 920–930 (2000).
157. Biebricher, C. K. & Orgel, L. E. An RNA that Multiplies Indefinitely with DNA-Dependent RNA Polymerase: Selection from a Random Copolymer. *Proc. Natl. Acad. Sci.* 70, 934–938 (1973).
158. Biebricher, C. K., Eigen, M. & Gardiner, W. C. Kinetics of RNA Replication: Competition and Selection Among Self-Replicating RNA Species. *Biochemistry (Mosc.)* 24, 6550–6560 (1985).
159. Robertson, D. L. & Joyce, G. F. Selection In Vitro of an RNA Enzyme that Specifically Cleaves Single-Stranded DNA. *Nature* 344, 467–468 (1990).
160. Beaudry, A. A. & Joyce, G. F. Directed Evolution of an RNA Enzyme. *Science* 257, 635–641 (1992).
161. Tuerk, C. & Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* 249, 505–510 (1990).
162. Ellington, A. D. & Szostak, J. W. In Vitro Selection of RNA Molecules that Bind Specific Ligands. *Nature* 346, 818–822 (1990).
163. Mills, D. R., Peterson, R. L. & Spiegelman, S. An Extracellular Darwinian Evolution Experiment with a Self-Duplicating Nucleic Acid Molecule. *Proc Natl Acad Sci USA* 58, 217–224 (1967).
164. Joyce, G. F. & Orgel, L. E. in *Cold Spring Harbor Monograph Series; The RNA world, Second edition* (eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 37, 1–25 (1993).
165. Sutherland, J. D. Ribonucleotides. *Cold Spring Harb. Perspect. Biol.* 2, (2010).
166. Shapiro, R. The Improbability of Prebiotic Nucleic Acid Synthesis. *Orig. Life* 14, 565–570 (1984).
167. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* 4, a003608 (2012).
168. Shapiro, R. A Simpler Origin for Life. *Sci. Am.* 296, 46–53 (2007).
169. Orgel, L. E. & Lohrmann, R. Prebiotic Chemistry and Nucleic Acid Replication. *Acc. Chem. Res.* 7, 368–377 (1974).
170. Butlerov, A. Académie des sciences (France). Comptes rendus hebdomadaires des séances de l'Académie des sciences (1861)

171. Breslow, R. On the Mechanism of the Formose Reaction. *Tetrahedron Lett.* 1, 22–26 (1959).
172. Schutte, W. A., Allamandola, L. J. & Sandford, S. A. An Experimental Study of the Organic Molecules Produced in Cometary and Interstellar Ice Analogs by Thermal Formaldehyde Reactions. *Icarus* 104, 118–137 (1993).
173. Fomenkova, M. N., Chang, S. & Mukhin, L. M. Carbonaceous Components in the Comet Halley Dust. *Geochim. Cosmochim. Acta* 58, 4503–4512 (1994).
174. Biver, N. *et al.* in *Cometary Science after Hale-Bopp* (eds. Boehnhardt, H., Combi, M., Kidger, M. R. & Schulz, R.) 323–333 (Springer Netherlands, 2002).
175. Bossard, A. R., Raulin, F., Mourey, D. & Toupance, G. Organic Synthesis from Reducing Models of the Atmosphere of the Primitive Earth with UV Light and Electric Discharges. *J. Mol. Evol.* 18, 173–178
176. Stribling, R. & Miller, S. L. Energy Yields for Hydrogen Cyanide and Formaldehyde Syntheses: The HCN and Amino Acid Concentrations in the Primitive Ocean. *Orig. Life Evol. Biosph.* 17, 261–273
177. Schlesinger, G. & Miller, S. L. Prebiotic Synthesis in Atmospheres Containing CH_4, CO, and CO_2. *J. Mol. Evol.* 19, 376–382
178. Cleaves II, H. J. The Prebiotic Geochemistry of Formaldehyde. *Precambrian Res.* 164, 111–118 (2008).
179. Socha, R. F., Weiss, A. H. & Sakharov, M. M. Autocatalysis in the Formose Reaction. *React. Kinet. Catal. Lett.* 14, 119–128 (1980).
180. Decker, P., Schweer, H. & Pohlamnn, R. Bioids : X. Identification of Formose Sugars, Presumable Prebiotic Metabolites, Using Capillary Gas Chromatography/Gas Chromatography—Mass Spectrometry of N-Butoxime Trifluoroacetates on OV-225. *J. Chromatogr. A* 244, 281–291 (1982).
181. Zubay, G. Studies on the Lead-Catalyzed Synthesis of Aldopentoses. *Orig. Life Evol. Biosph.* 28, 13–26 (1998).
182. Harsch, G., Bauer, H. & Voelter, W. Kinetik, Katalyse und Mechanismus der Sekundärreaktion in der Schlußphase der Formose-Reaktion. *Liebigs Ann. Chem.* 1984, 623–635 (1984).
183. Löb, W. Über das Verhalten des Formamids unter der Wirkung der stillen Entladung Ein Beitrag zur Frage der Stickstoff-Assimilation. *Berichte Dtsch. Chem. Ges.* 46, 684–697 (1913).
184. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate Minerals Stabilize Ribose. *Science* 303, 196–196 (2004).
185. Fernandez-Lopez, R., Kofoed, J., Machuqueiro, M. & Darbre, T. A Selective Direct Aldol Reaction in Aqueous Media Catalyzed by Zinc–Proline. *Eur. J. Org. Chem.* 2005, 5268–5276 (2005).

186. Kofoed, J., Reymond, J.-L. & Darbre, T. Prebiotic Carbohydrate Synthesis: Zinc–Proline Catalyzes Direct Aqueous Aldol Reactions of α-Hydroxy Aldehydes and Ketones. *Org. Biomol. Chem.* 3, 1850 (2005).
187. Mueller, D. *et al.* Chemistry of α-Aminonitriles. Aldomerization of Glycolaldehyde Phosphate to rac-Hexose 2,4,6-Triphosphates and (in Presence of Formaldehyde) rac-Pentose 2,4-Diphosphates: rac-Allose 2,4,6-Triphosphate and rac-Ribose 2,4-Diphosphate are the Main Reaction Products. *Helv. Chim. Acta* 73, 1410–68 (1990).
188. Larralde, R., Robertson, M. P. & Miller, S. L. Rates of Decomposition of Ribose and Other Sugars: Implications for Chemical Evolution. *Proc. Natl. Acad. Sci.* 92, 8158–8160 (1995).
189. Krishnamurthy, R., Guntha, S. & Eschenmoser, A. Regioselective α-Phosphorylation of Aldoses in Aqueous Solution. *Angew. Chem. Int. Ed.* 39, 2281–2285 (2000).
190. Shapiro, R. Prebiotic Ribose Synthesis: A Critical Analysis. *Orig. Life Evol. Biosph.* 18, 71–85 (1988).
191. Oró, J. Synthesis of Adenine from Ammonium Cyanide. *Biochem. Biophys. Res. Commun.* 2, 407–412 (1960).
192. Oró, J. & Kimball, A. P. Synthesis of Purines Under Possible Primitive Earth Conditions. I. Adenine from Hydrogen Cyanide. *Arch. Biochem. Biophys.* 94, 217–227 (1961).
193. Miller, S. L. The Mechanism of Synthesis of Amino Acids by Electric Discharges. *Biochim. Biophys. Acta* 23, 480–489 (1957).
194. Chyba, C. & Sagan, C. Endogenous Production, Exogenous Delivery and Impact-Shock Synthesis of Organic Molecules: an Inventory for the Origins of Life. *Nature* 355, 125–132 (1992).
195. Levine, J. S., Augustsson, T. R. & Natarajan, M. The Prebiological Paleoatmosphere: Stability and Composition. *Orig. Life* 12, 245–259 (1982).
196. Pizzarello, S. & Williams, L. B. Ammonia in the Early Solar System: An Account from Carbonaceous Meteorites. *Astrophys. J.* 749, 161 (2012).
197. Schwartz, A. W., Joosten, H. & Voet, A. B. Prebiotic Adenine Synthesis via HCN Oligomerization in Ice. *Biosystems* 15, 191–193 (1982).
198. Schwartz, A. W. & Goverde, M. Acceleration of HCN Oligomerization by Formaldehyde and Related Compounds: Implications for Prebiotic Syntheses. *J. Mol. Evol.* 18, 351–353 (1982).
199. Sanchez, R. A., Ferbis, J. P. & Orgel, L. E. Studies in Prebiotic Synthesis: II. Synthesis of purine precursors and amino acids from Aqueous Hydrogen Cyanide. *J. Mol. Biol.* 30, 223–253 (1967).
200. Ferris, J. P. & Orgel, L. E. An Unusual Photochemical Rearrangement in the Synthesis of Adenine from Hydrogen Cyanide1. *J. Am. Chem. Soc.* 88, 1074–1074 (1966).

201. Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in Prebiotic Synthesis: IV. Conversion of 4-aminoimidazole-5-carbonitrile Derivatives to Purines. *J. Mol. Biol.* 38, 121–128 (1968).
202. Albert, A. & Brown. Purine Studies. Part I. Stability to Acid and Alkali. Solubility. Ionization. Comparison with Pteridines.
203. Ferris, J. P., Sanchez, R. A. & Orgel, L. E. Studies In Prebiotic Synthesis: III. Synthesis of Pyrimidines from Cyanoacetylene and Cyanate. *J. Mol. Biol.* 33, 693–704 (1968).
204. Wen, N. & Brooker, M. H. Rate Constants for Cyanate Hydrolysis to Urea: A Raman Study. *Can. J. Chem.* 72, 1099–1106 (1994).
205. Ferris, J. P., Goldstein, G. & Beaulieu, D. J. Chemical Evolution. IV. Evaluation of Cyanovinyl Phosphate as a Prebiotic Phosphorylating Agent. *J. Am. Chem. Soc.* 92, 6598–6603 (1970).
206. Robertson, M. P. & Miller, S. L. An Efficient Prebiotic Synthesis of Cytosine and Uracil. *Nature* 375, 772–774 (1995).
207. Anastasi, C. *et al.* RNA: Prebiotic Product, or Biotic Invention? *Chem. Biodivers.* 4, 721–739 (2007).
208. Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in Prebiotic Synthesis: VI. Synthesis of Purine Nucleosides. *J. Mol. Biol.* 67, 25–33 (1972).
209. Drew, K. N., Zajicek, J., Bondo, G., Bose, B. & Serianni, A. S. 13C-Labeled Aldopentoses: Detection and Quantitation of Cyclic and Acyclic Forms by Heteronuclear 1D and 2D NMR Spectroscopy. *Carbohydr. Res.* 307, 199–209 (1998).
210. Praseuth, D. *et al.* Double Helices with Parallel Strands are Formed by Nuclease-Resistant Oligo-[A]-Deoxynucleotides and Oligo-[A]-Deoxynucleotides Covalently Linked to an Intercalating Agent with Complementary Oligo-[B]-Deoxynucleotides. *J. Mol. Biol.* 196, 939–942 (1987).
211. Ueda, T. & Nishino, H. On the Hilbert-Johnson Procedure for Pyrimidine Nucleoside Synthesis. *J. Am. Chem. Soc.* 90, 1678–1679 (1968).
212. Lazcano, A. & Miller, S. L. The Origin and Early Evolution of Life: Prebiotic Chemistry, the Pre-RNA World, and Time. *Cell* 85, 793–798 (1996).
213. Becker, S. *et al.* A High-Yielding, Strictly Regioselective Prebiotic Purine Nucleoside Formation Pathway. *Science* 352, 833–836 (2016).
214. Westheimer, F. H. Why Nature Chose Phosphates. *Science* 235, (1987).
215. Davis, B. D. On the Importance of Being Ionized. *Arch. Biochem. Biophys.* 78, 497–509 (1958).
216. Benner, S. A. & Hutter, D. Phosphates, DNA, and the Search for Nonterrean Life: A Second Generation Model for Genetic Molecules. *Bioorganic Chem.* 30, 62–80 (2002).

217. Pasek, M. A., Harnmeijer, J. P., Buick, R., Gull, M. & Atlas, Z. Evidence for Reactive Reduced Phosphorus Species in the Early Archean Ocean. *Proc. Natl. Acad. Sci.* 110, 10089–10094 (2013).
218. Yamagata, Y., Watanabe, H., Saitoh, M. & Namba, T. Volcanic Production of Polyphosphates and its Relevance to Prebiotic Evolution. *Nature* 352, 516–519 (1991).
219. Bryant, D. E. *et al.* On the Prebiotic Potential of Reduced Oxidation State Phosphorus: the H-Phosphinate–Pyruvate System. *Chem. Commun.* 46, 3726–3728 (2010).
220. Pasek, M. A., Kee, T. P., Bryant, D. E., Pavlov, A. A. & Lunine, J. I. Production of Potentially Prebiotic Condensed Phosphates by Phosphorus Redox Chemistry. *Angew. Chem. Int. Ed.* 47, 7918–7920 (2008).
221. Pasek, M. A. & Lauretta, D. S. Aqueous Corrosion of Phosphide Minerals from Iron Meteorites: A Highly Reactive Source of Prebiotic Phosphorus on the Surface of the Early Earth. *Astrobiology* 5, 515–535 (2005).
222. Schwartz, A. W. Prebiotic Phosphorus Chemistry Reconsidered. *Orig. Life Evol. Biosph.* 27, 505–512
223. Lohrmann, R. & Orgel, L. E. Prebiotic Synthesis: Phosphorylation in Aqueous Solution. *Science* 161, 64–66 (1968).
224. Saffhill, R. Selective Phosphorylation of the cis-2',3'-diol of Unprotected Ribonucleosides with Trimetaphosphate in Aqueous Solution. *J. Org. Chem.* 35, 2881–2883 (1970).
225. Lohrmann, R. & Orgel, L. E. Urea-Inorganic Phosphate Mixtures as Prebiotic Phosphorylating Agents. *Science* 171, 490–494 (1971).
226. Schoffstall, A. M. Prebiotic Phosphorylation of Nucleosides in Formamide. *Orig. Life* 7, 399–412 (1976).
227. Beck, A., Lohrmann, R. & Orgel, L. E. Phosphorylation with Inorganic Phosphates at Moderate Temperatures. *Science* 157, 952–952 (1967).
228. Handschuh, G. J., Lohrmann, R. & Orgel, L. E. The Effect of Mg2+ and Ca2+ on Urea-Catalyzed Phosphorylation Reactions. *J. Mol. Evol.* 2, 251–262 (1973).
229. Reimann, R. & Zubay, G. Nucleoside Phosphorylation: A Feasible Step in the Prebiotic Pathway to RNA. *Orig. Life Evol. Biosph.* 29, 229–247 (1999).
230. Krishnamurthy, R., Arrhenius, G. & Eschenmoser, A. Formation of Glycolaldehyde Phosphate from Glycolaldehyde in Aqueous Solution. *Orig. Life Evol. Biosph.* 29, 333–354 (1999).
231. Halmann, M., Sanchez, R. A. & Orgel, L. E. Phosphorylation of D-Ribose in Aqueous Solution. *J. Org. Chem.* 34, 3702–3703 (1969).
232. Renz, M., Lohrmann, R. & Orgel, L. E. Catalysts for the polymerization of adenosine cyclic 2′,3′-phosphate on a poly (U) template. *Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth.* 240, 463–471 (1971).

233. Usher, D. A. & McHale, A. H. Nonenzymic Joining of Oligoadenylates on a Polyuridylic Acid Template. *Science* 192, 53–54 (1976).
234. Bolli, M., Micura, R., Pitsch, S. & Eschenmoser, A. Pyranosyl-RNA: Further Observations on Replication. *Helv. Chim. Acta* 80, 1901–1951 (1997).
235. Bowler, F. R. *et al.* Prebiotically Plausible Oligoribonucleotide Ligation Facilitated by Chemoselective Acetylation. *Nat. Chem.* 5, 383–389 (2013).
236. Crowe, M. A. & Sutherland, J. D. Reaction of Cytidine Nucleotides with Cyanoacetylene: Support for the Intermediacy of Nucleoside-2′,3′-cyclic Phosphates in the Prebiotic Synthesis of RNA. *ChemBioChem* 7, 951–956 (2006).
237. Mullen, L. B. & Sutherland, J. D. Simultaneous Nucleotide Activation and Synthesis of Amino Acid Amides by a Potentially Prebiotic Multi-Component Reaction. *Angew. Chem. Int. Ed.* 46, 8063–8066 (2007).
238. Kanavarioti, A., Monnard, P.-A. & Deamer, D. W. Eutectic Phases in Ice Facilitate Nonenzymatic Nucleic Acid Synthesis. *Astrobiology* 1, 271–281 (2001).
239. Deck, C., Jauker, M. & Richert, C. Efficient Enzyme-Free Copying of all Four Nucleobases Templated by Immobilized RNA. *Nat. Chem.* 3, 603–608 (2011).
240. Prabahar, K. J. & Ferris, J. P. Adenine Derivatives as Phosphate-Activating Groups for the Regioselective Formation of 3',5'-Linked Oligoadenylates on Montmorillonite: Possible Phosphate-Activating Groups for the Prebiotic Synthesis of RNA. *J. Am. Chem. Soc.* 119, 4330–4337 (1997).
241. Ferris, J. P., Joshi, P. C., Wang, K.-J., Miyakawa, S. & Huang, W. Catalysis in Prebiotic Chemistry: Application to the Synthesis of RNA Oligomers. *Adv. Space Res.* 33, 100–105 (2004).
242. Lohrmann, R. Formation of Nucleoside 5′-Phosphoramidates Under Potentially Prebiological Conditions. *J. Mol. Evol.* 10, 137–154
243. Rohatgi, R., Bartel, D. P. & Szostak, J. W. Nonenzymatic, Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3'−5' Phosphodiester Bonds. *J. Am. Chem. Soc.* 118, 3340–3344 (1996).
244. Inoue, T. & Orgel, L. E. Oligomerization of (Guanosine 5′-phosphor)-2-methylimidazolide on Poly(C). *J. Mol. Biol.* 162, 201–217 (1982).
245. Inoue, T. *et al.* Template-Directed Synthesis on the Pentanucleotide CpCpGpCpC. *J. Mol. Biol.* 178, 669–676 (1984).
246. Yamagata, Y. Prebiotic Formation of ADP and ATP from AMP, Calcium Phosphates and Cyanate in Aqueous Solution. *Orig. Life Evol. Biosph.* 29, 511–520
247. Rohatgi, R., Bartel, D. P. & Szostak, J. W. Kinetic and Mechanistic Analysis of Nonenzymatic, Template-Directed Oligoribonucleotide Ligation. *J. Am. Chem. Soc.* 118, 3332–3339 (1996).

248. Ferris, J. P. Mineral Catalysis and Prebiotic Synthesis: Montmorillonite-Catalyzed Formation of RNA. *Elements* 1, 145–149 (2005).
249. Huang, W. & Ferris, J. P. One-Step, Regioselective Synthesis of up to 50-mers of RNA Oligomers by Montmorillonite Catalysis. *J. Am. Chem. Soc.* 128, 8914–8919 (2006).
250. Wu, T. & Orgel, L. E. Nonenzymic Template-Directed synthesis on Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides. *J. Am. Chem. Soc.* 114, 317–322 (1992).
251. Landweber, L. F. & Pokrovskaya, I. D. Emergence of a Dual-Catalytic RNA with Metal-Specific Cleavage and Ligase Activities: The Spandrels of RNA Evolution. *Proc. Natl. Acad. Sci.* 96, 173–178 (1999).
252. Rogers, J. & Joyce, G. F. A Ribozyme that lacks Cytidine. *Nature* 402, 323–325 (1999).
253. Attwater, J., Wochner, A. & Holliger, P. In-Ice Evolution of RNA Polymerase Ribozyme Activity. *Nat. Chem.* 5, 1011–1018 (2013).
254. Kauffman, S. A. *The Origins of Order: Self Organization and Selection in Evolution*. *Oxford University Press* (1993).
255. Lincoln, T. A. & Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. *Science* 323, 1229–1232 (2009).
256. Pitsch, S. *et al.* Pyranosyl-RNA ('p-RNA'): Base-Pairing Selectivity and Potential to Replicate. Preliminary Communication. *Helv. Chim. Acta* 78, 1621–1635 (1995).
257. Eschenmoser, A. Towards a Chemical Etiology of Nucleic Acid Structure. *Orig. Life Evol. Biosph.* 27, 535–553 (1997).
258. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. Peptide Nucleic Acids (PNA). Oligonucleotide Analogs with an Achiral Peptide Backbone. *J. Am. Chem. Soc.* 114, 1895–1897 (1992).
259. Egholm, M. *et al.* PNA Hybridizes to Complementary Oligonucleotides Obeying the Watson–Crick Hydrogen-Bonding Rules. *Nature* 365, 566–568 (1993).
260. Schmidt, J. G., Christensen, L., Nielsen, P. E. & Orgel, L. E. Information Transfer from DNA to Peptide Nucleic Acids by Template-Directed Syntheses. *Nucleic Acids Res.* 25, 4792–4796 (1997).
261. Schöning, K.-U. *et al.* Chemical Etiology of Nucleic Acid Structure: The α-Threofuranosyl-(3'→2') Oligonucleotide System. *Science* 290, 1347–1351 (2000).
262. Herdewijn, P. TNA as a Potential Alternative to Natural Nucleic Acids. *Angew. Chem. Int. Ed.* 40, 2249–2251 (2001).
263. Zhang, L., Peritz, A. & Meggers, E. A Simple Glycol Nucleic Acid. *J. Am. Chem. Soc.* 127, 4174–4175 (2005).

264. Diederichsen, U. Alaynl PNA: Evidence for Linear Band Structures Based on Guanine-Cytosin Base Pairs. *Angew. Chem. Int. Ed. Engl.* 36, 1886–1889 (1997).
265. Diederichsen, U. Pairing Properties of Alanyl Peptide Nucleic Acids Containing an Amino Acid Backbone with Alternating Configuration. *Angew. Chem. Int. Ed. Engl.* 35, 445–448 (1996).
266. Cairns-Smith, A. G. The Origin of Life and the Nature of the Primitive Gene. *J. Theor. Biol.* 10, 53–88 (1966).
267. Weber, A. L. The Triose Model: Glyceraldehyde as a Source of Energy and Monomers for Prebiotic Condensation Reactions. *Orig. Life Evol. Biosph.* 17, 107–119 (1987).
268. Powner, M. W. & Sutherland, J. D. Prebiotic Chemistry: a New Modus Operandi. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 2870–2877 (2011).
269. Kiedrowski, G. von, Otto, S. & Herdewijn, P. Welcome Home, Systems Chemists! *J. Syst. Chem.* 1, 1 (2010).
270. Sanchez, R. A. & Orgel, L. E. Studies in Prebiotic Synthesis: V. Synthesis and Photoanomerization of Pyrimidine Nucleosides. *J. Mol. Biol.* 47, 531–543 (1970).
271. Tapiero, C. M. & Nagyvary, J. Prebiotic Formation of Cytidine Nucleotides. *Nature* 231, 42–43 (1971).
272. Anastasi, C., Crowe, M. A., Powner, M. W. & Sutherland, J. D. Direct Assembly of Nucleoside Precursors from Two- and Three-Carbon Units. *Angew. Chem.* 118, 6322–6325 (2006).
273. Müller, D. *et al.* Chemie von a-Aminonitrilen. Aldomerisierung von Glycolaldehyd-phosphat zu racemischen Hexose-2,4,6-triphosphaten und (in Gegenwart von Formaldehyd) racemischen Pentose-2,4-diphosphaten: rac-Allose-2,4,6-triphosphat und rac-Ribose-2,4-diphosphat sind die Reaktionshauptprodukte. *Helv. Chim. Acta* 73, 1410–1468 (1990).
274. Powner, M. W., Sutherland, J. D. & Szostak, J. W. Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. *J. Am. Chem. Soc.* 132, 16677–16688 (2010).
275. Ritson, D. & Sutherland, J. D. Prebiotic Synthesis of Simple Sugars by Photoredox Systems Chemistry. *Nat. Chem.* 4, 895–899 (2012).
276. Ritson, D. J. & Sutherland, J. D. Synthesis of Aldehydic Ribonucleotide and Amino Acid Precursors by Photoredox Chemistry. *Angew. Chem. Int. Ed.* 52, 5845–5847 (2013).
277. Donn, B. Comets: Chemistry and Chemical Evolution. *J. Mol. Evol.* 18, 157–160
278. Hanel, R. *et al.* Infrared Observations of the Saturnian System from Voyager 1. *Science* 212, 192–200 (1981).
279. Kurosawa, K., Sugita, S., Ishibashi, K., Hasegawa, S., Sekine, Y., Ogawa, N. O., & Matsui, T. .Hydrogen Cyanide Production due to Mid-Size Impacts in

a Redox-Neutral N2-Rich Atmosphere. *Origins of Life and Evolution of Biospheres*, *43*(3), 221-245 (2013).
280. Strecker, A. Ueber Einen Neuen aus Aldehyd - Ammoniak und Blausäure entstehenden Körper. *Justus Liebigs Ann. Chem.* 91, 349–351 (1854).
281. Miller, S. L. & Urey, H. C. Organic Compound Synthesis on the Primitive Earth. *Science* 130, 245–251 (1959).
282. Moutou, G. *et al.* Equilibrium of α-Aminoacetonitrile Formation from Formaldehyde, Hydrogen Cyanide and Ammonia in Aqueous Solution: Industrial and Prebiotic Significance. *J. Phys. Org. Chem.* 8, 721–730 (1995).
283. Anastasi, C., Crowe, M. A. & Sutherland, J. D. Two-Step Potentially Prebiotic Synthesis of α-d-Cytidine-5'-phosphate from D-Glyceraldehyde-3-phosphate. *J. Am. Chem. Soc.* 129, 24–25 (2007).
284. Ii, H. J. C. The Prebiotic Synthesis of Acrolein. *Monatshefte Für Chem. Chem. Mon.* 134, 585–593 (2003).
285. Corma, A., Huber, G. W., Sauvanaud, L. & O'Connor, P. Biomass to chemicals: Catalytic Conversion of Glycerol/Water Mixtures into Acrolein, Reaction Network. *J. Catal.* 257, 163–171 (2008).
286. Tsukuda, E., Sato, S., Takahashi, R. & Sodesawa, T. Production of Acrolein from Glycerol Over Silica-Supported Heteropoly Acids. *Catal. Commun.* 8, 1349–1353 (2007).
287. Payne, G. B. Reactions of Hydrogen Peroxide. V.1 Alkaline Epoxidation of Acrolein and Methacrolein. *J. Am. Chem. Soc.* 81, 4901–4904 (1959).
288. Hurd, C. D. & Gershbein, L. L. Reactions of Mercaptans with Acrylic and Methacrylic Derivatives. *J. Am. Chem. Soc.* 69, 2328–2335 (1947).
289. Islam, S., Bucar, K. & Powner, M. W. Prebiotic Selection and Assembly of Proteinogenic Amino Acids and Natural Nucleotides from Complex Mixtures. *Nat. Chem.* (accepted, *awaiting publication*), (2016).
290. Coggins, A. J. & Powner, M. W. Prebiotic Synthesis of Phosphoenol Pyruvate by α-Phosphorylation-Controlled Triose Glycolysis. *Nat. Chem.*, (2016).
291. Walsh, C. T., Benson, T. E., Kim, D. H. & Lees, W. J. The Versatility of Phosphoenolpyruvate and its Vinyl Ether Products in Biosynthesis. *Chem. Biol.* 3, 83–91 (1996).
292. Kasting, J. F., Holland, H. D. & Pinto, J. P. Oxidant abundances in Rainwater and the Evolution of Atmospheric Oxygen. *J. Geophys. Res. Atmospheres* 90, 10497–10510 (1985).
293. Borda, M. J., Elsetinow, A. R., Strongin, D. R. & Schoonen, M. A. A Mechanism for the Production of Hydroxyl Radical at Surface Defect Sites on Pyrite. *Geochim. Cosmochim. Acta* 67, 935–939 (2003).
294. Borda, M. J., Elsetinow, A. R., Schoonen, M. A. & Strongin, D. R. Pyrite-Induced Hydrogen Peroxide Formation as a Driving Force in the Evolution

of Photosynthetic Organisms on an Early Earth. *Astrobiology* 1, 283–288 (2001).
295. Cuppen, H. M., Ioppolo, S., Romanzin, C. & Linnartz, H. Water formation at Low Temperatures by Surface O2 Hydrogenation II: the Reaction Network. *Phys. Chem. Chem. Phys.* 12, 12077–12088 (2010).
296. Foustoukos, D. I., Houghton, J. L., Seyfried Jr., W. E., Sievert, S. M. & Cody, G. D. Kinetics of H2–O2–H2O Redox Equilibria and Formation of Metastable H2O2 under Low Temperature Hydrothermal Conditions. *Geochim. Cosmochim. Acta* 75, 1594–1607 (2011).
297. Nagorski, R. W. & Richard, J. P. Mechanistic Imperatives for Aldose-Ketose Isomerization in Water: Specific, General Base- and Metal Ion-Catalyzed Isomerization of Glyceraldehyde with Proton and Hydride Transfer. *J Am Chem Soc* 123 (5), 794–802 (2001).
298. Appayee, C. & Breslow, R. Deuterium Studies Reveal a New Mechanism for the Formose Reaction Involving Hydride Shifts. *J. Am. Chem. Soc.* 136, 3720–3723 (2014).
299. Choudhary, A., Kamer, K. J., Powner, M. W., Sutherland, J. D. & Raines, R. T. A Stereoelectronic Effect in Prebiotic Nucleotide Synthesis. *ACS Chem. Biol.* 5, 655–657 (2010).
300. Saladino, R. *et al.* From Formamide to RNA: the Roles of Formamide and Water in the Evolution of Chemical Information. *Res. Microbiol.* 160, 441–448 (2009).
301. Pitsch, S., Pombo-Villar, E. & Eschenmoser, A. Chemie von α-Aminonitrilen. 13. Mitteilung. Über die Bildung von 2-Oxoethyl-phosphaten ('Glycoladehyd-phosphaten') aus rac-Oxirancarbonitril und anorganischem Phosphat und über (formale) Konstitutionelle Zusammenhänge zwischen 2-Oxoethyl-phosphaten und Oligo (hexo- und pentopyranosyl)nucleotid-Rückgraten. *Helv. Chim. Acta* 77, 2251–2285 (1994).
302. Krishnamurthy, R., Pitsch, S. & Arrhenius, G. Mineral Induced Formation of Pentose-2,4-Bisphosphates. *Orig. Life Evol. Biosph.* 29, 139–152
303. Fringuelli, F., Pizzo, F. & Vaccaro, L. NaOH-Catalyzed Thiolysis of α,β-Epoxyketones in Water. A Key Step in the Synthesis of Target Molecules Starting from α,β-Unsaturated Ketones. *J. Org. Chem.* 69, 2315–2321 (2004).
304. Gauss, D., Schoenenberger, B. & Wohlgemuth, R. Chemical and Enzymatic Methodologies for the Synthesis of Enantiomerically Pure Glyceraldehyde 3-Phosphates. *Carbohydr. Res.* 389, 18–24 (2014).
305. Springsteen, G. & Joyce, G. F. Selective Derivatization and Sequestration of Ribose from a Prebiotic Mix. *J. Am. Chem. Soc.* 126, 9578–9583 (2004).
306. Saewan, N. *et al.* Exploratory Studies to Investigate a Linked Prebiotic Origin of RNA and Coded Peptides. 4th Communication. *Chem. Biodivers.* 2, 66–83 (2005).

307. Tenud, L., Farooq, S., Seibl, J. & Eschenmoser, A. Endocyclische SN-Reaktionen am gesättigten Kohlenstoff? Vorläufige Mitteilung. *Helv. Chim. Acta* 53, 2059–2069 (1970).
308. Powner, M. Studies Towards a Chemical Origin of RNA. *PhD thesis, University of Manchester* (2009).
309. Powner, M. W. & Sutherland, J. D. Phosphate-Mediated Interconversion of Ribo- and Arabino-Configured Prebiotic Nucleotide Intermediates. *Angew. Chem.* 122, 4745–4747 (2010).
310. Cockerill, A. F. *et al.* An Improved Synthesis of 2-Amino-1,3-oxazoles under Basic Catalysis. *Synthesis* 1976, 591–593 (1976).
311. Powner, M. W., Zheng, S.-L. & Szostak, J. W. Multicomponent Assembly of Proposed DNA Precursors in Water. *J. Am. Chem. Soc.* 134, 13889–13895 (2012).
312. Shannahoff, D. H. & Sanchez, R. A. 2,2'-Anhydropyrimidine nucleosides. Novel Syntheses and Reactions. *J. Org. Chem.* 38, 593–598 (1973).
313. Beck, A., Lohrmann, R. & Orgel, L. E. Phosphorylation with Inorganic Phosphates at Moderate Temperatures. *Science* 157, 952–952 (1967).
314. Ferris, J. P., Goldstein, G. & Beaulieu, D. J. Chemical evolution. IV. Evaluation of Cyanovinyl Phosphate as a Prebiotic Phosphorylating Agent. *J. Am. Chem. Soc.* 92, 6598–6603 (1970).
315. Choudhary, A., Kamer, K. J., Powner, M. W., Sutherland, J. D. & Raines, R. T. A Stereoelectronic Effect in Prebiotic Nucleotide Synthesis. *ACS Chem. Biol.* 5, 655–657 (2010).
316. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature* 459, 239–242 (2009).
317. Engelhart, A. E., Powner, M. W. & Szostak, J. W. Functional RNAs exhibit tolerance for non-heritable 2′–5′ versus 3′–5′ backbone heterogeneity. *Nat. Chem.* 5, 390–394 (2013).
318. Eftink, M. R. & Biltonen, R. L. Energetics of ribonuclease A catalysis. 2. Nonenzymatic hydrolysis of cytidine cyclic 2',3'-phosphate. *Biochemistry (Mosc.)* 22, 5134–5140 (1983).
319. Bowler, F. R. *et al.* Prebiotically Plausible Oligoribonucleotide Ligation Facilitated by Chemoselective Acetylation. *Nat. Chem.* 5, 383–389 (2013).
320. Verlander, M. S., Lohrmann, R. & Orgel, L. E. Catalysts for the Self-Polymerization of Adenosine Cyclic 2′,3′-phosphate. *J. Mol. Evol.* 2, 303–316 (1973).
321. Ferris, J. P., Hill, A. R., Liu, R. & Orgel, L. E. Synthesis of Long Prebiotic Oligomers on Mineral Surfaces. *Nature* 381, 59–61 (1996).
322. *Unpublished results of ongoing work in the Powner lab*, performed by Dr. C. A. Fernandez.

323. Powner, M. W. *et al.* On the Prebiotic Synthesis of Ribonucleotides: Photoanomerisation of Cytosine Nucleosides and Nucleotides Revisited. *ChemBioChem* 8, 1170–1179 (2007).
324. Powner, M. W. & Sutherland, J. D. Potentially Prebiotic Synthesis of Pyrimidine β-D-Ribonucleotides by Photoanomerization/Hydrolysis of α-D-Cytidine-2′-Phosphate. *ChemBioChem* 9, 2386–2387 (2008).
325. Coulter, C. L. Structural Chemistry of Cyclic Nucleotides. II. Crystal and Molecular Structure of Sodium β-Cytidine 2',3'-cyclic Phosphate. *J. Am. Chem. Soc.* 95, 570–575 (1973).
326. Rammler, D. H., Lapidot, Y. & Khorana, H. G. Studies on Polynucleotides. XIX.1 The Specific Synthesis of C3″-C5″ Inter-Ribonucleotidic Linkage.2 A New Approach and its Use in the Synthesis of C3″-C5″-Linked Uridine Oligonucleotides3. *J. Am. Chem. Soc.* 85, 1989–1997 (1963).
327. Biron, J.-P., Parkes, A. L., Pascal, R. & Sutherland, J. D. Expeditious, Potentially Primordial, Aminoacylation of Nucleotides. *Angew. Chem. Int. Ed.* 44, 6731–6734 (2005).
328. Liu, R. & Orgel, L. E. Oxidative Acylation using Thioacids. *Nature* 389, 52–54 (1997).
329. Nagyvary, J. Arabinonucleotides. II. Synthesis of O2,2'-anhydrocytidine 3'-Phosphate, a Precursor of 1-β-D-arabinosylcytosine. *J. Am. Chem. Soc.* 91, 5409–5410 (1969).
330. Eschenmoser, A. On a Hypothetical Generational Relationship between HCN and Constituents of the Reductive Citric Acid Cycle. *Chem. Biodivers.* 4, 554–573 (2007).
331. Kofoed, J., Darbre, T. & Reymond, J.-L. Dual Mechanism of Zinc-Proline Catalyzed Aldol Reactions in Water - Chemical Communications (RSC Publishing). *Chem. Commun.*
332. Pitsch, S., Eschenmoser, A., Gedulin, B., Hui, S. & Arrhenius, G. Mineral Induced Formation of Sugar Phosphates. *Orig. Life Evol. Biosph.* 25, 297–334 (1995).
333. Powner, M., Sutherland, J. & Szostak, J. The Origins of Nucleotides. *Synlett* 2011, 1956–1964 (2011).
334. Polazzi, J. O., Leland, D. L. & Kotick, M. P. Rearrangment of Anhydropyrimidine Nucleosides in Liquid Hydrogen Fluoride. Mechanism, Scope, and Synthetic Studies. *J. Org. Chem.* 39, 3114–3119 (1974).
335. Mukaiyama, T., Hayashi, Y. & Hashimoto, Y. Highly Stereoselective Synthesis of α-Ribonucleotides via a Cyclic Phosphate. *Chem. Lett.* 1087–1090 (1985).
336. Mizuno, Y. & Sasaki, T. A New 'Anhydronucleoside Method' for the Synthesis of Dinucleoside Phosphates. *J. Am. Chem. Soc.* 88, 863–864 (1966).

337. Paquette, L. A., Zeng, Q., Tsui, H.-C. & Johnston, J. N. Stereochemical Models for the Enantiocontrolled Construction of Fully Functionalized C Rings via Intramolecular Aldolization in Advanced Precursors to Paclitaxel. *J. Org. Chem.* 63, 8491–8509 (1998).
338. Angelin, M., Hermansson, M., Dong, H. & Ramström, O. Direct, Mild, and Selective Synthesis of Unprotected Dialdo-Glycosides. *Eur. J. Org. Chem.* 2006, 4323–4326 (2006).
339. Golding, B. T., Slaich, P. K., Kennedy, G., Bleasdale, C. & Watson, W. P. Mechanisms of Formation of Adducts from Reactions of Glycidaldehyde with 2'-Deoxyguanosine and/or Guanosine. *Chem. Res. Toxicol.* 9, 147–157 (1996).
340. Grelbig, T., Pötter, B. & Seppelt, K. Die Stärke von Schwefel-Kohlenstoff-Mehrfachbindungen. *Chem. Ber.* 120, 815–817 (1987).
341. Trevino, S. G., Zhang, N., Elenko, M. P., Lupták, A. & Szostak, J. W. Evolution of Functional Nucleic Acids in the Presence of Nonheritable Backbone Heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13492–13497 (2011).
342. Alberty, R. A. & Goldberg, R. N. Standard Thermodynamic Formation Properties for the Adenosine 5'-Triphosphate Series. *Biochemistry (Mosc.)* 31, 10610–10615 (1992).
343. Pasek, M. A. Rethinking Early Earth Phosphorus Geochemistry. *Proc. Natl. Acad. Sci.* 105, 853–858 (2008).
344. Pasek, M. A., Dworkin, J. P. & Lauretta, D. S. A radical pathway for Organic Phosphorylation During Schreibersite Corrosion with Implications for the Origin Of Life. *Geochim. Cosmochim. Acta* 71, 1721–1736 (2007).
345. Peyser, J. R. & Ferris, J. P. The Rates of Hydrolysis of Thymidyl-3′,5′-Thymidine-H-Phosphonate: The Possible Role of Nucleic Acids Linked by Diesters of Phosphorous Acid in the Origins of Life. *Orig. Life Evol. Biosph.* 31, 363–380
346. Poopeiko, N. E., Juhl, M., Vester, B., Sørensen, M. D. & Wengel, J. Xylo-Configured Oligonucleotides (XNA, Xylo Nucleic Acid): Synthesis of Conformationally Restricted Derivatives and Hybridization Towards DNA and RNA Complements. *Bioorg. Med. Chem. Lett.* 13, 2285–2290 (2003).
347. Schöppe, A., Hinz, H.-J., Rosemeyer, H. & Seela, F. Xylose-DNA: Comparison of the Thermodynamic Stability of Oligo(2′-deoxyxylonucleotide) and Oligo(2′-deoxyribonucleotide) Duplexes. *Eur. J. Biochem.* 239, 33–41 (1996).
348. Juliá, S., Masana, J. & Vega, J. C. 'Synthetic Enzymes'. Highly Stereoselective Epoxidation of Chalcone in a Triphasic Toluene-Water-Poly[(S)-alanine] System. *Angew. Chem. Int. Ed. Engl.* 19, 929–931 (1980).

349. Shaffer, C. L., Morton, M. D. & Hanzlik, R. P. N-Dealkylation of an N-Cyclopropylamine by Horseradish Peroxidase. Fate of the Cyclopropyl Group. *J. Am. Chem. Soc.* 123, 8502–8508 (2001).
350. Nemet, I., Vikić-Topić, D. & Varga-Defterdarović, L. Spectroscopic Studies of Methylglyoxal in Water and Dimethylsulfoxide. *Bioorganic Chem.* 32, 560–570 (2004).
351. Serianni, A. S., Pierce, J. & Barker, R. Carbon-13-Enriched Carbohydrates: Preparation of Triose, Tetrose, and Pentose Phosphates. *Biochemistry (Mosc.)* 18, 1192–1199 (1979).
352. Ahn, M., Cochrane, F. C., Patchett, M. L. & Parker, E. J. Arabinose 5-Phosphate Analogues as Mechanistic Probes for Neisseria Meningitidis 3-Deoxy-D-Manno-Octulosonate 8-Phosphate Synthase. *Bioorg. Med. Chem.* 16, 9830–9836 (2008).
353. Tsai, Y.-F. *et al.* Asymmetric Synthesis of (-)-Muricatacin's Analogue (4S,5R)-5-Hydroxy-4-octadecanolide Exhibiting the Cytotoxicity against Esphageal Cancer Cells. *Heterocycles* 85, 299 (2012).
354. Islam, S. *et al.* Detection of Potential TNA and RNA Nucleoside Precursors in a Prebiotic Mixture by Pure Shift Diffusion-Ordered NMR Spectroscopy. *Chem. – Eur. J.* 19, 4586–4595 (2013).
355. Powner, M. W. & Sutherland, J. D. Phosphate-Mediated Interconversion of Ribo- and Arabino-Configured Prebiotic Nucleotide Intermediates. *Angew. Chem. Int. Ed.* 49, 4641–4643 (2010).

11. Appendix

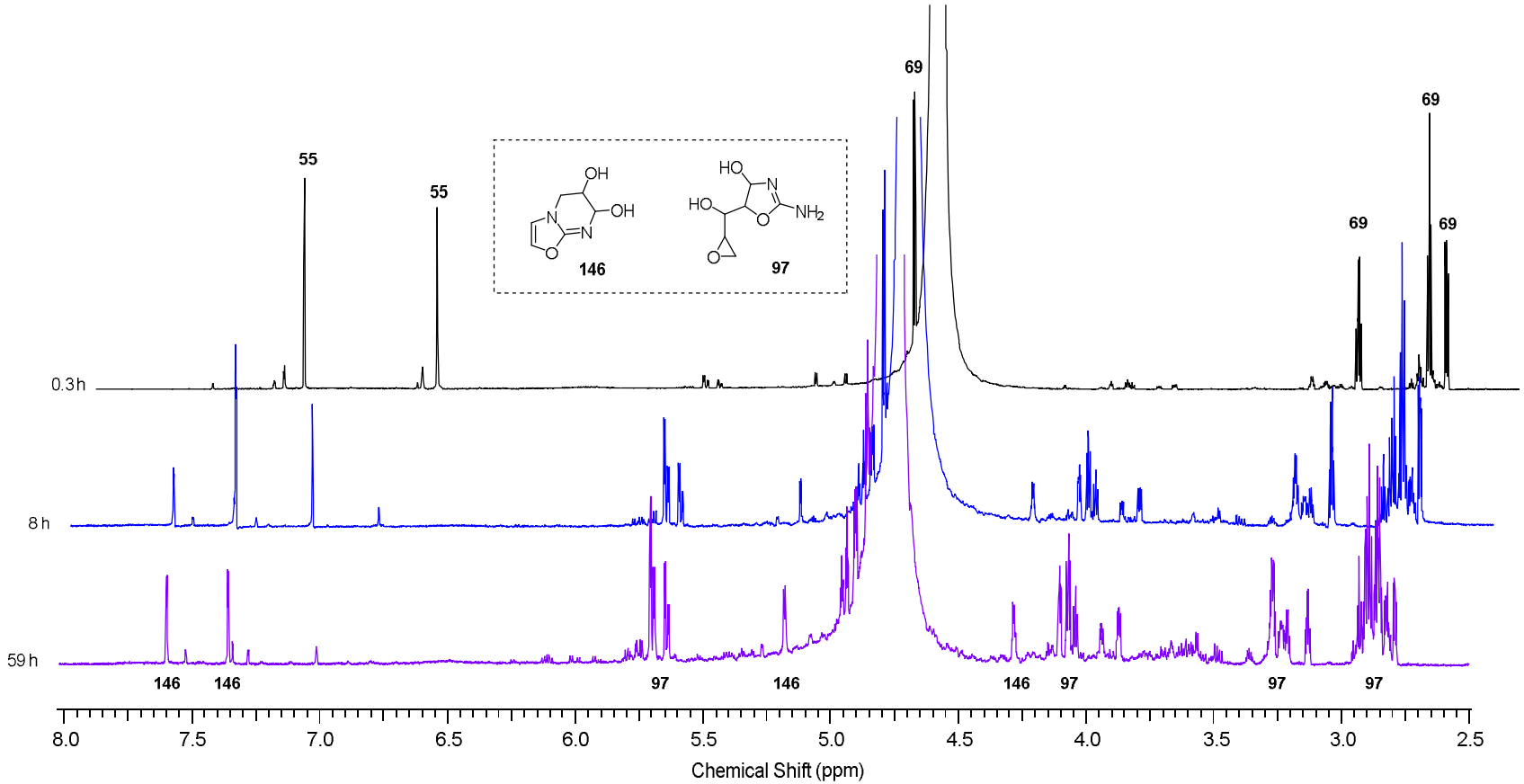


*Figure 11.1: [1]H NMR spectra (600 MHz, H_2O/D_2O 9:1, 8.00-2.50 ppm) showing the consumption of glycidaldehyde (**69**) over time to form adduct **97** (and minor amount of by-product **146**). Reaction of **69** (128mM) and 2-aminooxazole (**55**; 1.1 eq) in water at pH 5, RT. For ease of analysis, only the most downfield proton peaks of **97** (between 5.55-5.75 ppm) are used in analysis. Offset 0.25 ppm.*

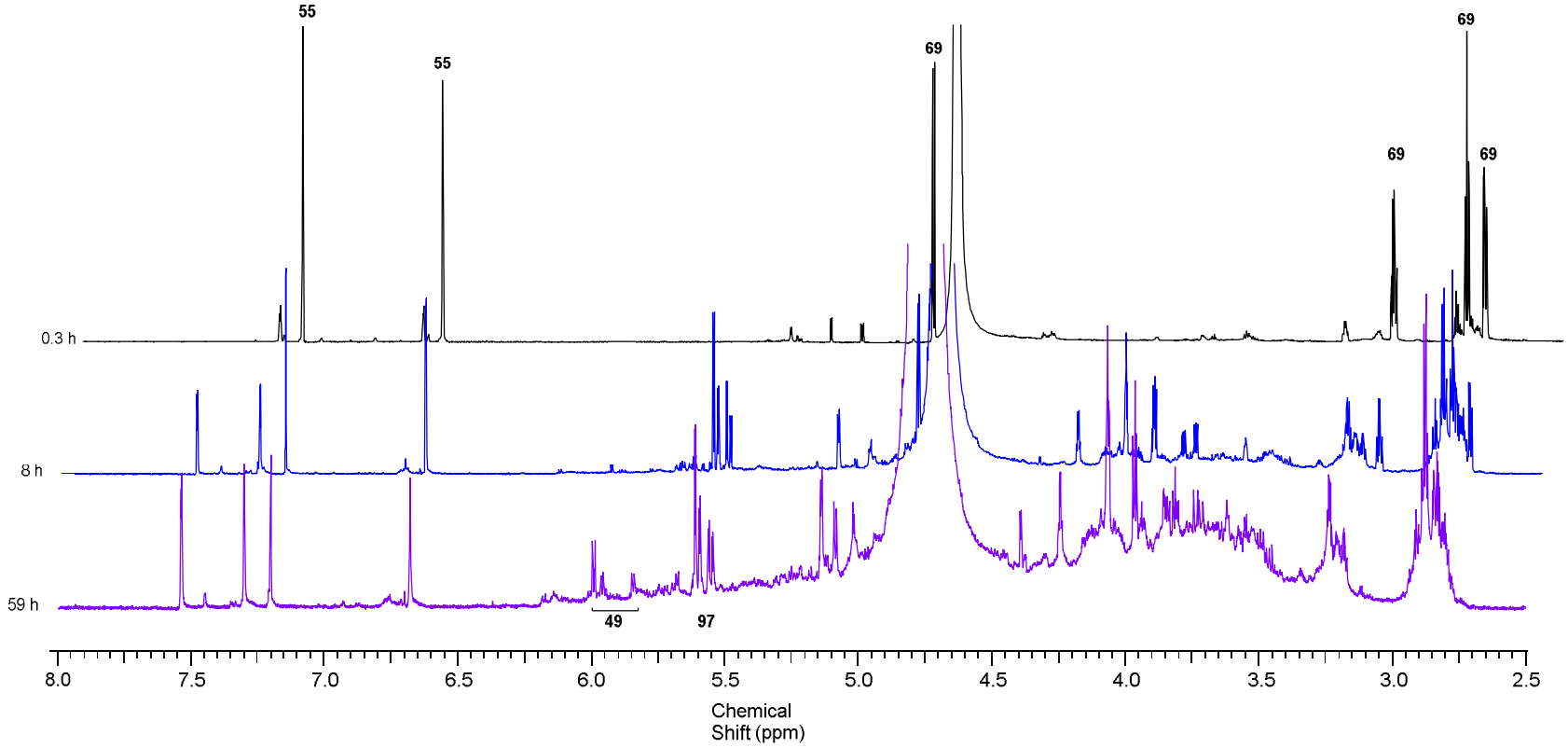


*Figure 11.2: [1]H NMR spectra (600 MHz, H_2O/D_2O 9:1, 8.00-2.50 ppm) showing the consumption of glycidaldehyde (**69**) over time to form adduct **97** and aminooxazolines (**49**). Reaction of **69** (128mM) and 2-aminooxazole (**55**; 1.1 eq) in water at pH 7, RT. For ease of analysis, only the most downfield proton peaks of **97** (between 5.55-5.75 ppm) are annotated and used in analysis. Offset 0.25 ppm.*

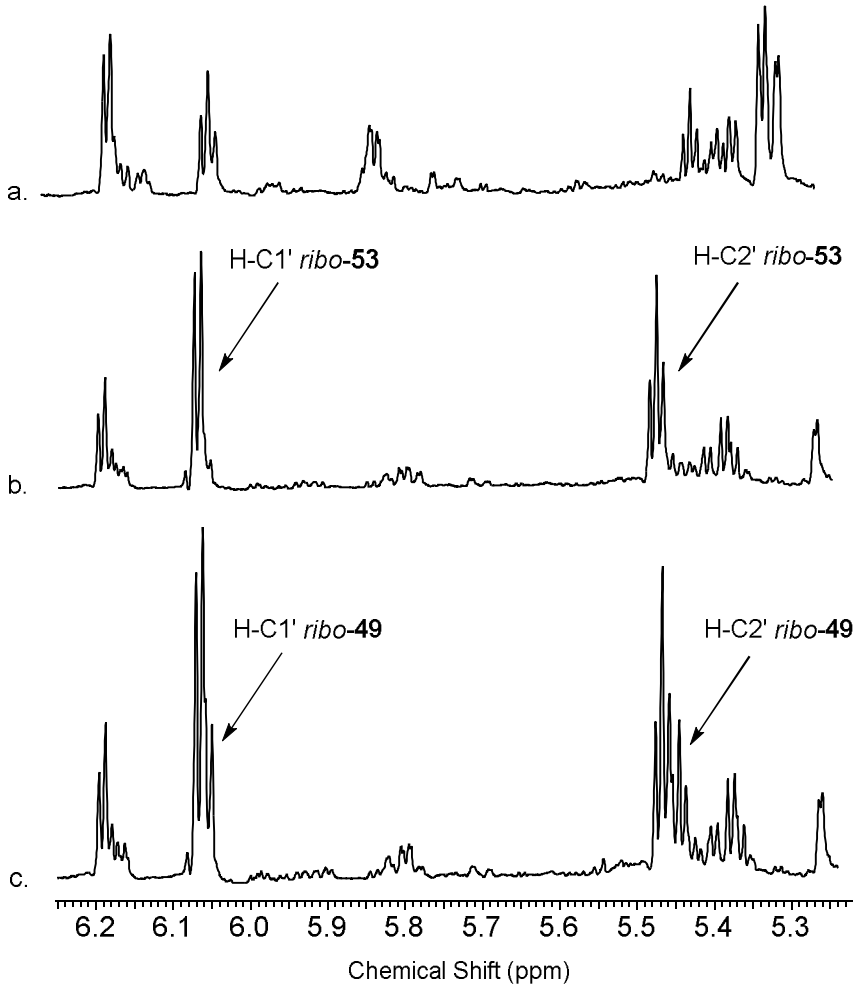


*Figure 11.3: ^{1}H NMR spectra (600 MHz, D_2O, 6.25-5.23 ppm) showing the H-C1' and H-C2' peak region for aminooxazolines, and subsequent signal amplification observed by spiking with a synthetically prepared authentic standard. Reaction of adduct **97** with P_i (3 eq) at pH 7. a. Reaction was changed to pH 5 after running 28 days at pH 7, in the hope of separating overlapping peaks in the 6 ppm region. b. Spiked with riboaminooxazoline 5'-phosphate (ribo-**53**). c. Spiked with riboaminooxazoline (ribo-**49**).*

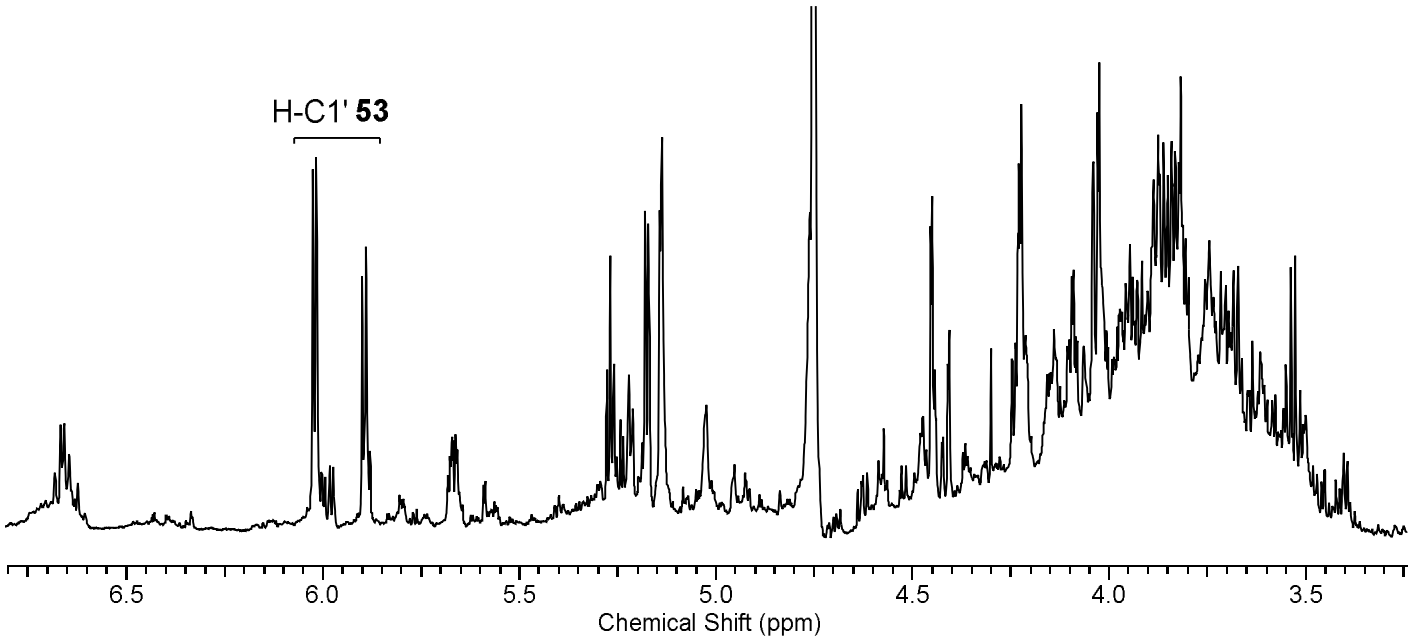


*Figure 11.4: ^{1}H NMR spectrum (600 MHz, H_2O/D_2O 9:1, 6.80-3.25 ppm) showing the formation of 5'-phosphorylated aminooxazolines (**53**) from glycidaldehyde (**69**; 128mM), 2-aminooxazole (**55**; 1.1 eq) and P_i (3 eq) after 14 days at pH 7.*

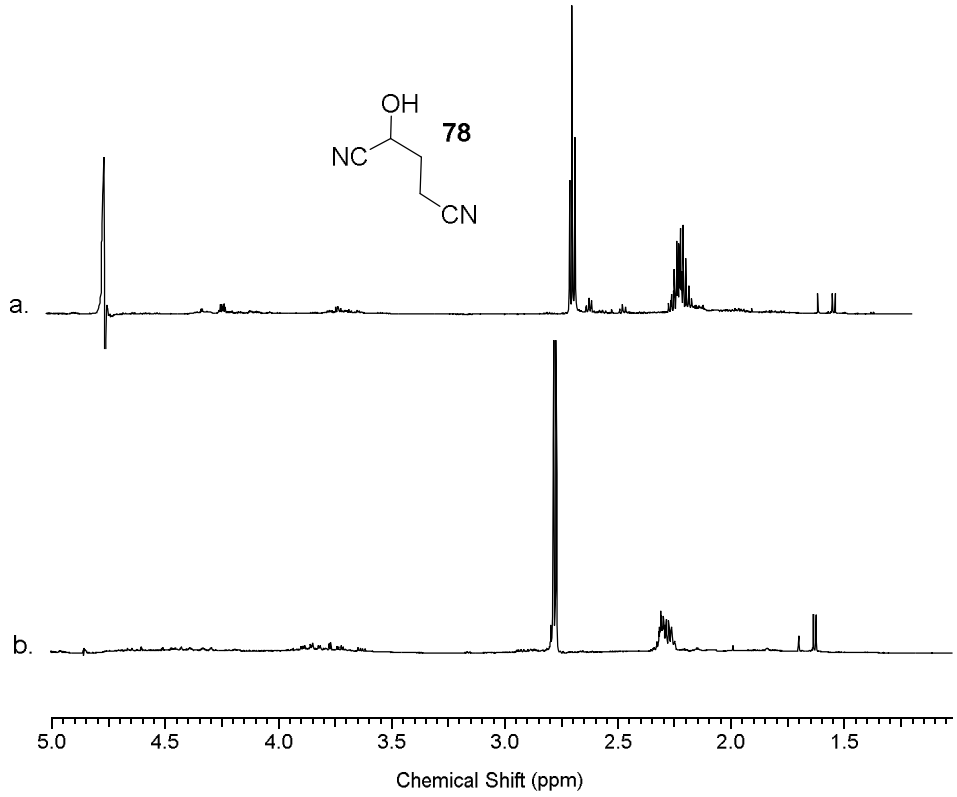


*Figure 11.5: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 9.00-1.00 ppm) a. Reaction of acrolein (**1**; 140mM) with potassium cyanide (5 eq) at pH 9 after 24 h at RT, showing the synthesis of 2-hydroxypentanedinitrile (**78**); b. ^{1}H NMR (600 MHz, D_2O) Same reaction as* a. *done in D_2O (pD 10, 24 h at RT).*

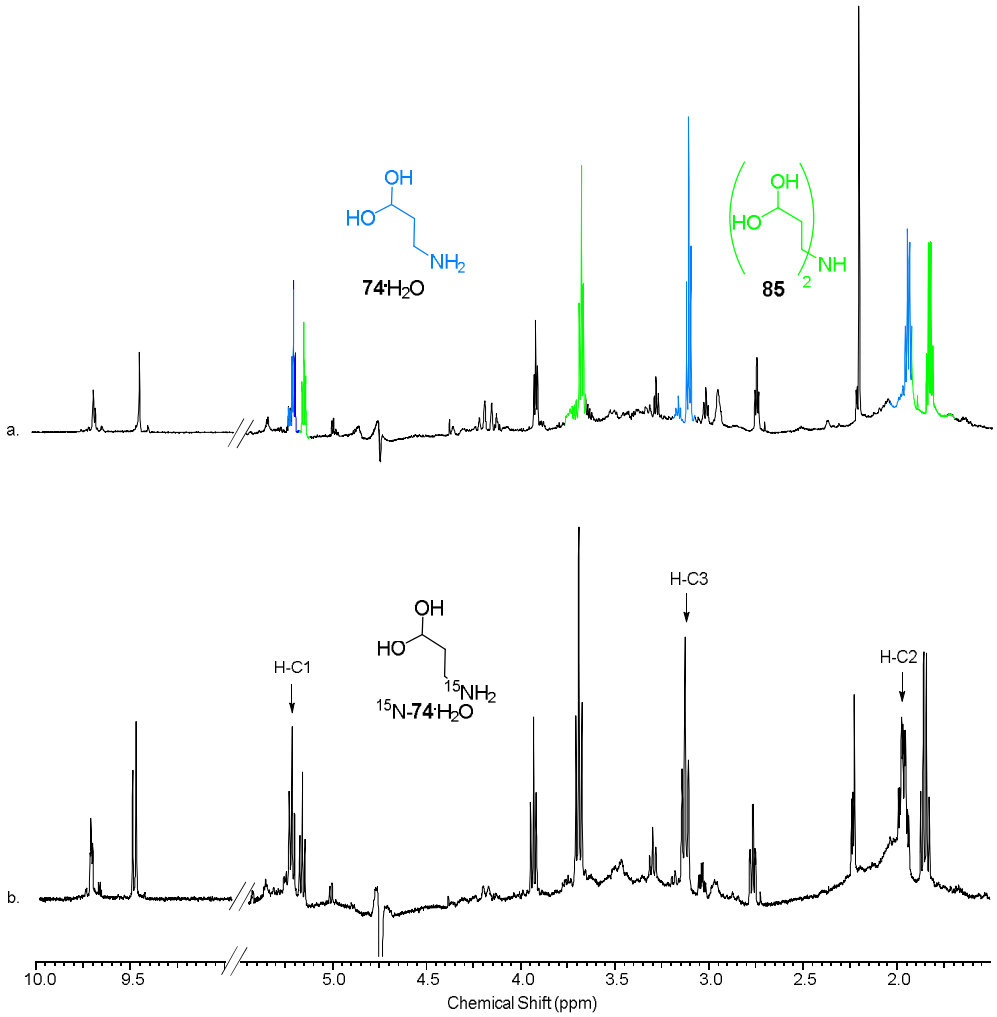


*Figure 11.6: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 10.00-1.50 ppm) a. Reaction of acrolein (**1**) with NH_4OH (1:5 **1**:NH_4OH) at pH 6.5 after 12 h at RT, showing the synthesis of 3-aminopropanal hydrate (**74**·H_2O); b. Reaction of **1** with $^{15}NH_4OH$ (1:5 **1**: $^{15}NH_4OH$) at pH 6.5 after 48 h at RT, showing the synthesis of ^{15}N-**74**·H_2O.*

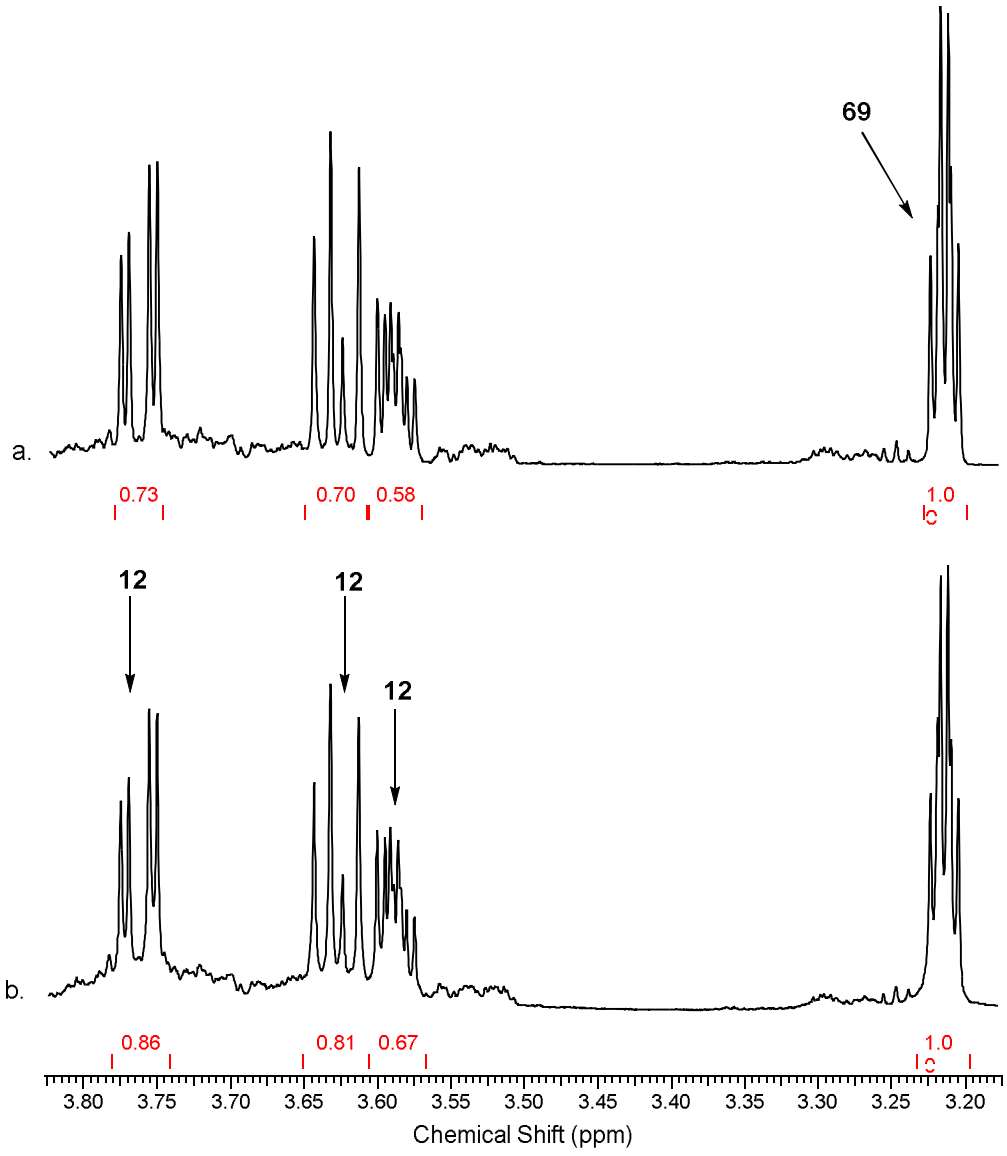


*Figure 11.7: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 3.82-3.18 ppm) showing signal amplification observed by spiking with a commercial standard. a. Hydrolysis experiment of glycidaldehyde (**69**; 0.4M, pH 4) yielding glyceraldehyde (**12**) after 120 h; b. spiked with an authentic sample of **12**.*

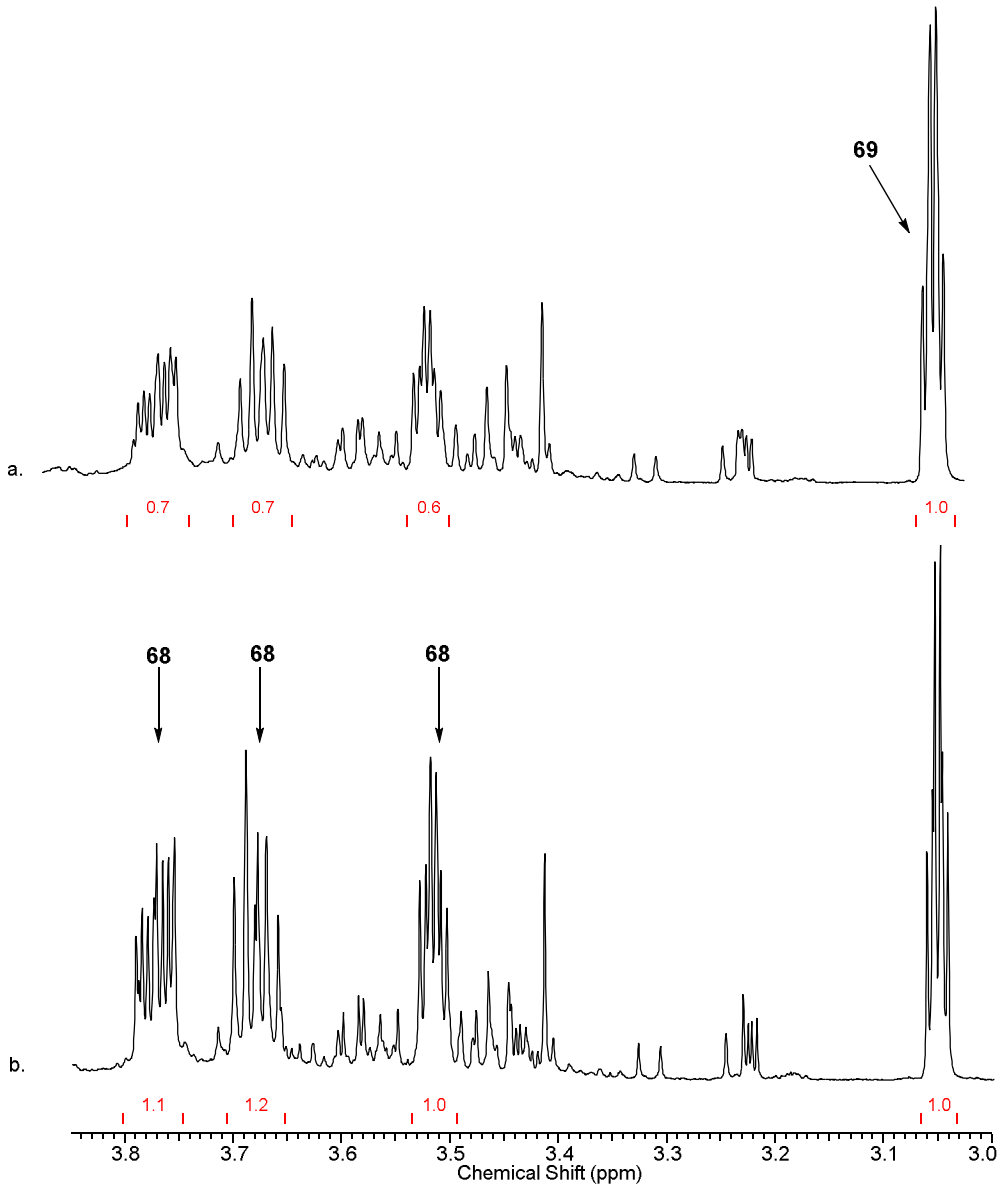


*Figure 11.8: ^{1}H NMR spectra (600 MHz, H_2O/D_2O 9:1, 3.85-3.00 ppm) of a. the reaction of glycidaldehyde (**69**; 0.25M) in P_i buffer (4 eq; pH 7) after incubation at RT for 92 h. b. spiked with a commercial sample of glyceraldehyde 3-phosphate (**68**).*